La saga du web multilingue

Par Marie Lebert, 7 décembre 2019.

LE LIVRE DANS LES COMMUNITY TEXTS DE L'INTERNET ARCHIVE

Suite à l'invention du web en 1990, les usagers de l'internet dont la langue maternelle n'est pas l'anglais atteignent les 5% en 1994, les 20% en 1998, les 50% en 2000 et les 75% en 2015. Nombreux sont ceux qui aident à promouvoir leur propre langue et culture et d'autres langues et cultures — parfois sur leur temps libre et souvent en utilisant l'anglais comme langue passerelle — pour que le web devienne vraiment multilingue. Ce livre basé sur de nombreux entretiens est un hommage à leur travail et leur ténacité.

[Version anglaise]
[Version espagnole]

---

* Prologue
* L'internet
* L'encodage
* L'internationalisation
* Le multilinguisme
* La localisation des sites web
* La collaboration en ligne
* Les textes électroniques
* Les livres numériques
* La presse
* La précarisation de l'emploi
* Le droit d'auteur
* Le droit d'auteur adapté au web
* Les librairies
* Les librairies numériques
* Les auteurs
* Les bibliothèques
* Les bibliothécaires
* Les bibliothèques numériques
* Les trésors du passé
* Les catalogues collectifs

* Les ressources linguistiques
* Les dictionnaires
* Les encyclopédies
* Les revues scientifiques
* Les ressources pour enseigner
* Les ressources pour les traducteurs
* Les bases terminologiques
* La traduction automatique
* La traduction assistée par ordinateur
* La traduction automatique gratuite en ligne
* Le catalogue des langues vivantes
* Les langues minoritaires
* Les langues menacées
* Chronologie

Prologue

Suite à l'invention du web en 1990, les usagers de l'internet dont la langue maternelle n'est pas l'anglais atteignent les 5% en 1994, les 20% en 1998, les 50% en 2000 et les 75% en 2015.

Brian King, directeur du WorldWide Language Institute (Institut des langues du monde entier), explique en septembre 1998 dans un entretien par courriel: «Dans un rapport de l'UNESCO du début des années 1950, l'enseignement dispensé dans sa langue maternelle était considéré comme un droit fondamental de l'enfant. La possibilité de naviguer sur l'internet dans sa langue maternelle pourrait bien être son équivalent à l' ge de l'Information. Si l'internet doit vraiment devenir le réseau mondial qu'on nous promet, tous les usagers devraient y avoir accès sans problème de langue. Le considérer comme la chasse gardée de ceux qui, par accident historique, nécessité pratique ou privilège politique, connaissent l'anglais, est injuste à l'égard de ceux qui ne connaissent pas cette langue.»

Maria Victoria Marinetti, professeur d'espagnol en entreprise et traductrice, écrit en août 1999: «Il est très important de pouvoir communiquer en différentes langues sur l'internet. Je dirais même que c'est obligatoire, car l'information donnée sur l'internet est à destination du monde entier, alors pourquoi ne l'aurions-nous pas dans notre propre langue ou dans la langue que nous souhaitons utiliser? Information mondiale, mais pas de vaste choix dans les langues, ce serait contradictoire, pas vrai?»

Le nombre d'usagers vivant hors des États-Unis atteint les 50% en juillet 1999. Jean-Pierre Cloutier, auteur des Chroniques de Cybérie, une chronique hebdomadaire des actualités de l'internet, écrit en août 1999: «Cet été, le cap a été franchi. Plus de 50% des utilisateurs et utilisatrices du réseau sont hors des États-Unis. L'an prochain, plus de 50% des utilisateurs seront non anglophones. Il y a seulement cinq ans, c'était 5%. Formidable, non?»

Le nombre d’usagers dont la langue maternelle n’est pas l’anglais atteint bien les 50% en été 2000 et continue d’augmenter ensuite. Selon la société marketing Global Reach, ce pourcentage est de 52,5% en juillet 2001, 57% en décembre 2001, 59,8% en avril 2002 et 64,4% en septembre 2003.

Quinze ans après la création du web, le magazine Wired constate dans son numéro d’août 2005 que «moins de la moitié du web est commercial, le reste fonctionne avec la passion.» Quant à l’internet, d’après le quotidien Le Monde du 19 août 2005, «ses trois pouvoirs — l’ubiquité, la variété et l’interactivité — rendent son potentiel d’usages quasi infini.»

Nombreux sont ceux qui aident à promouvoir leur propre langue et culture et d’autres langues et cultures — parfois sur leur temps libre et souvent en utilisant l’anglais comme langue passerelle — pour que le web devienne vraiment multilingue. Ce livre basé sur de nombreux entretiens est un hommage à leur travail et leur ténacité.

L’internet

Henri “Henk” Slettenhaar, spécialiste des systèmes de communication, a vécu toutes les étapes de l’internet lors d’une longue carrière à Genève et en Californie. Européen trilingue, il a passé son enfance aux Pays-Bas, il dispense ses cours en anglais et il parle français puisqu’il vit en France.

En 1958, il rejoint le CERN (Organisation européenne pour la recherche nucléaire) à Genève pour travailler sur le premier ordinateur numérique et participer au développement des premiers réseaux numériques. Son expérience californienne débute en 1966 lorsqu’il rejoint pendant 18 mois une équipe du SLAC (Stanford Linear Accelerator Center — Centre de l’accélérateur linéaire de Stanford) pour créer un numérisateur de film. De retour au SLAC en 1983, il conçoit un système numérique de contrôle qui sera utilisé pendant dix ans. Il est ensuite professeur en technologies des communications à la Webster University de Genève pendant 25 ans, et devient en 2000 le directeur de son Telecom Management Program (Programme de gestion des télécoms). En 1992, il crée la Silicon Valley Association (SVA), une association suisse qui organise des voyages d’étude dans la Silicon Valley et dans d’autres pôles de haute technologie.

Henri Slettenhaar raconte en décembre 1998 dans un entretien par courriel: «Je ne peux pas imaginer ma vie professionnelle sans l’internet. Cela fait vingt ans que j’utilise le courrier électronique. Les premières années, c’était le plus souvent pour communiquer avec mes collègues dans un secteur géographique très limité. Depuis l’explosion de l’internet et l’avènement du web, je communique principalement par courriel, mes conférences sont en grande partie sur le web et mes cours ont tous un prolongement sur le web. Pour les visites que j’organise dans la Silicon Valley, toutes les informations sont disponibles sur le web, et je ne pourrais pas organiser ces visites sans utiliser l’internet. De plus, l’internet est pour moi une fantastique base de données disponible en quelques clics de souris.»

Il écrit aussi: «Les communautés locales présentes sur le web devraient en tout premier lieu utiliser leur langue pour diffuser des informations. Si elles veulent présenter ces informations à la communauté internationale, celles-ci doivent être disponibles aussi en anglais. Je pense qu'il existe un réel besoin de sites bilingues. Mais je suis également enchanté qu'il existe maintenant autant de documents disponibles dans leur langue originale. Je préfère de beaucoup lire l'original avec difficulté plutôt qu'une traduction médiocre.»

Il ajoute en août 1999 : «À mon avis, il existe deux catégories de sites web. La première catégorie vise une recherche globale dans le domaine des affaires et de l'information. Pour cela, la langue est d'abord l'anglais, avec des versions locales si nécessaire. La seconde catégorie consiste en des informations locales de tous ordres dans les endroits les plus reculés. Si l'information est à destination d'une ethnie ou d'un groupe linguistique, elle doit d'abord être dans la langue de l'ethnie ou du groupe linguistique, avec peut-être un résumé en anglais.»

Il ajoute en août 2000: «Le multilinguisme s'est beaucoup développé. De nombreux sites de commerce électronique sont devenus multilingues, et il existe maintenant des sociétés qui vendent des produits permettant la localisation des sites web. Je note aussi l'explosion de la technologie du mobile. Le téléphone mobile est devenu pour beaucoup de gens, moi y compris, le moyen de communication personnel vous permettant d'être joignable à tout moment où que vous soyez. Toutefois l'internet mobile est encore du domaine du rêve. Les nouveaux services offerts par les téléphones GSM sont extrêmement primitifs et très chers, si bien que le WAP a reçu le sobriquet de "Wait And Pay" [Attendez et payez].»

Il ajoute en juillet 2001: «Ce qui me vient à l'esprit est le changement considérable apporté par le fait que j'ai maintenant une connexion à débit rapide chez moi. Le fait d'être constamment connecté est totalement différent du fait de se connecter de temps à autre par le biais de la ligne téléphonique. Je reçois maintenant mes messages dès leur arrivée dans ma messagerie. Je peux écouter mes stations radio préférées où qu'elles soient dans le monde. Je peux écouter les actualités quand je veux. Je peux également écouter la musique que j'aime à longueur de journée. La seule chose qui manque est une vidéo de bonne qualité en temps réel. La largeur de bande passante est encore insuffisante pour cela. Mon domicile est maintenant équipé d'un LAN [réseau internet local] avec et sans fil. Je peux utiliser mon ordinateur portable partout à l'intérieur et à l'extérieur de la maison, et même chez les voisins, tout en restant connecté. La même technologie me permet maintenant d'utiliser la carte LAN de mon ordinateur lorsque je voyage. Par exemple, lors de mon dernier voyage à Stockholm, je pouvais être connecté à l'internet à l'hôtel, au centre de conférences, à l'aéroport et même au pub irlandais!»

Pierre Schweitzer, concepteur du baladeur de textes @folio, écrit pour sa part en décembre 2006: «La chance qu'on a tous est de vivre là, ici et maintenant cette transformation fantastique. Quand je suis né en 1963, les ordinateurs avaient comme mémoire quelques pages de caractères à peine. Aujourd'hui, mon baladeur de musique pourrait contenir des milliards de pages, une vraie bibliothèque de quartier. Demain, par l'effet conjugué de la loi de Moore et de l'omniprésence des réseaux, l'accès instantané aux oeuvres et aux savoirs sera de mise. Le support de stockage lui-même n'aura plus beaucoup d'intérêt. Seules importeront les commodités fonctionnelles d'usage et la poétique de ces objets.»

Jean-Paul, auteur hypermédia et webmestre du site Cotres.net, écrit en janvier 2007: «J’ai l’impression que nous vivons une période “flottante”, entre les temps héroïques, où il s’agissait d’avancer en attendant que la technologie nous rattrape, et le futur, où le très haut débit va libérer les forces qui commencent à bouger, pour l’instant dans les seuls jeux. Les deux grandes tendances des prochaines années seront (a) la généralisation du très haut débit (c’est-à-dire en fait du débit normal), qui va m’affranchir des limitations purement techniques, notamment des soucis de poids et d’affichage des fichiers (mort définitive, enfin, des histogrammes de chargement); (b) le développement de la 3D. C’est le récit en hypermédia (= le multimédia + le clic) qui m’intéresse. Les pièges que pose un récit en 2D sont déjà passionnants. Avec la 3D, il va falloir chevaucher le tigre pour éviter la simple prouesse technique et laisser la priorité au récit.»

Comment l’infrastructure de l’internet va-t-elle évoluer? D’après Rafi Haladjian, fondateur de la société technologique Ozone, l’internet pourrait devenir un réseau pervasif permettant de se connecter en tout lieu et à tout moment sur tout type d’appareil à travers un réseau unique et omniprésent. Il explique en 2007 sur le site web de sa société: «La nouvelle vague touchera notre monde physique, notre environnement réel, notre vie quotidienne dans tous les instants. Nous n’accéderons plus au réseau, nous l’habiterons. Les composantes futures de ce réseau (parties filaires, parties non filaires, opérateurs) seront transparentes à l’utilisateur final. Il sera toujours ouvert, assurant une permanence de la connexion en tout lieu. Il sera également agnostique en terme d’application(s), puisque fondé sur les protocoles mêmes de l’internet.»

Quant au contenu du réseau pervasif, il pourrait être celui décrit dès 1994 par le philosophe Timothy Leary dans son livre «Chaos et cyberculture»: «Toute l’information du monde est à l’intérieur. Et grâce au cyberespace, tout le monde peut y avoir accès. Tous les signaux humains contenus jusque-là dans les livres ont été numérisés. Ils sont enregistrés et disponibles dans ces banques de données, sans compter tous les tableaux, tous les films, toutes les émissions de télé, tout, absolument tout.»

L’encodage

Le premier système d’encodage informatique est l’ASCII (American Standard Code for Information Interchange — Code standard américain pour les échanges d’information). Publié en 1963 par l’American National Standards Institute (ANSI — Institut américain pour les standards nationaux), l’ASCII est un code standard de 128 caractères convertis en langage binaire sur sept bits (A devient «1000001», B devient par «1000010», etc.). Les 128 caractères comprennent 33 caractères de contrôle (qui ne représentent donc pas de symbole écrit) et 95 caractères imprimables: les 26 lettres sans accent en majuscules (A-Z) et minuscules (a-z), les chiffres, les signes de ponctuation et quelques symboles, le tout correspondant aux touches du clavier anglophone QWERTY (différent du clavier français AZERTY).

L’ASCII permet uniquement la lecture de l’anglais (et du latin), et ne permet donc pas de prendre en compte les lettres accentuées présentes dans bon nombre de langues (français, espagnol, allemand, etc.), et à plus forte raison les langues utilisant d’autres alphabets (grec, arabe, russe, etc.) et les langues non alphabétiques (chinois, japonais, coréen, etc.). Ceci ne pose pas de problème majeur les premières années, tant que l’échange de fichiers électroniques se limite au public anglophone, à savoir 95% des usagers de l’internet au début des années 1990.

Suite à l’invention du web en 1990 et au lancement du navigateur Mosaic en décembre 1993, l’internet se développe hors de la sphère anglophone. Des variantes de l’ASCII sur huit bits sont créées pour traiter 256 caractères et prendre en compte les caractères accentués de quelques langues. La variante pour le français, l’espagnol et l’allemand (entre autres) est la norme ISO 8859-1, appelée aussi ISO Latin-1.

Puis l’échange des données s’internationalise encore davantage. Même avec les différentes variantes de l’ASCII, la situation devient compliquée, et le passage de l’ASCII original sur sept bits à ses variantes sur huit bits devient un véritable casse-tête, les problèmes étant entre autres la corruption des données dans les échanges informatiques ou encore l’incompatibilité des variantes entre elles, les pages ne pouvant être affichées que dans une seule langue à la fois.

Comme expliqué en septembre 1998 par Brian King, directeur du WorldWide Language Institute (Institut des langues du monde entier), lors d’un entretien par courriel: «La technologie des ordinateurs a longtemps été le seul domaine d’une élite “technicienne”, à l’aise à la fois dans des langages de programmation complexes et en anglais, la langue universelle des sciences et techniques. Les premiers ordinateurs n’ont jamais été conçus pour manier des systèmes d’écriture ne pouvant pas être convertis en ASCII. Il n’y avait pas de place pour autre chose que les 26 lettres de l’alphabet anglais dans un système d’encodage qui, à l’origine, ne pouvait même pas reconnaître les accents aigus et les trémas, sans parler des systèmes non alphabétiques comme le chinois. Mais la tradition a été bouleversée et la technologie popularisée. Des interfaces graphiques tels que ceux de Windows et de Macintosh [Apple] ont accéléré le processus. La stratégie de marketing de Microsoft a consisté à présenter son système d’exploitation comme facile à utiliser par le client moyen. À l’heure actuelle, cette facilité d’utilisation s’est étendue bien au-delà du PC, vers le réseau internet, si bien que, maintenant, même ceux qui ne sont pas programmeurs peuvent insérer des applets Java dans leurs pages web sans comprendre une seule ligne de programmation.»

«L’extension de cette popularisation locale est l’exportation des technologies de l’information dans le monde entier. La popularisation est maintenant effective à l’échelon mondial, et l’anglais n’est plus nécessairement la langue obligée de l’utilisateur. Il n’y a plus vraiment de langue indispensable, mais seulement les langues personnelles des utilisateurs. Une chose est certaine: il n’est plus nécessaire de comprendre l’anglais pour utiliser un ordinateur, de même qu’il n’est plus nécessaire d’avoir un diplôme d’informatique. La demande des utilisateurs non anglophones et l’effort entrepris par les sociétés technologiques se faisant concurrence pour obtenir les marchés mondiaux font que la localisation des sites est un secteur en expansion rapide pour le développement des logiciels et du matériel informatique. Le premier pas a été le passage de l’ASCII à l’ASCII étendu. Ceci signifie que les ordinateurs commençaient à

reconnaître les accents et les symboles utilisés dans les variantes de l’alphabet anglais, symboles qui appartenaient le plus souvent aux langues européennes. Cependant une page ne pouvait être affichée que dans une seule langue à la fois.»

«L’innovation la plus récente est Unicode. Bien qu’il soit encore en train d’évoluer et qu’il ait tout juste été incorporé [en septembre 1998] dans les derniers logiciels, ce nouveau système d’encodage convertit chaque caractère sur 16 bits. Alors que l’ASCII étendu à huit bits pouvait prendre en compte un maximum de 256 caractères, l’Unicode peut prendre en compte plus de 65.000 caractères uniques et il a donc la possibilité de traiter informatiquement tous les systèmes d’écriture de la planète. Les instruments sont maintenant plus ou moins en place. Ils ne sont pas encore parfaits, mais on peut désormais naviguer sur le web en chinois, en japonais, en coréen, et dans de nombreuses autres langues qui n’utilisent pas l’alphabet occidental. Comme l’internet s’étend à des parties du monde où l’anglais est très peu utilisé, par exemple la Chine, il est naturel que ce soit le chinois et non l’anglais qui soit utilisé. La majorité des usagers en Chine n’a pas d’autre choix que sa langue maternelle.»

Publié pour la première fois en janvier 1991, l’Unicode commence donc à se généraliser en 1998, par exemple pour les fichiers texte sous plateforme Windows (Windows NT, Windows 2000, Windows XP et versions suivantes), qui étaient jusque-là en ASCII. Mais l’utilisation de l’Unicode ne peut être que progressive. Un énorme travail est en effet nécessaire pour passer d’un système d’encodage à l’autre, avec des variantes UTF-8, UTF-16 et UTF-32 (UTF: Unicode Transformation Format) selon le nombre de bits utilisés. Il faudra attendre fin 2007 pour que l’Unicode supplante définitivement l’ASCII sur l’internet.

Patrick Rebollar, professeur de littérature française au Japon et modérateur de la liste de diffusion LITOR (Littérature et ordinateur), explique en janvier 2000: «Il s’agit d’abord d’un problème logiciel. Comme on le voit avec Netscape ou Internet Explorer, la possibilité d’affichage multilingue existe. La compatibilité entre ces logiciels et les autres (de la suite Office de Microsoft, par exemple) n’est cependant pas acquise. L’adoption de la table Unicode devrait résoudre une grande partie des problèmes, mais il faut pour cela réécrire la plupart des logiciels, ce à quoi les producteurs de logiciels rechignent du fait de la dépense, pour une rentabilité qui n’est pas évidente car ces logiciels entièrement multilingues intéressent moins de clients que les logiciels de navigation.»

Luc Dall’Armellina, co-auteur et webmestre de l’espace d’écriture hypermédia oVosite, écrit en juin 2000: «Les systèmes d’exploitation se dotent peu à peu des kits de langues et bientôt peut-être de polices de caractères Unicode à même de représenter toutes les langues du monde; reste que chaque application, du traitement de texte au navigateur web, emboîte ce pas. Les difficultés sont immenses: notre clavier avec ses ± 250 touches avoue ses manques dès lors qu’il faille saisir des Katakana ou Hiragana japonais, pire encore avec la langue chinoise. La grande variété des systèmes d’écriture de par le monde et le nombre de leurs signes font barrage. Mais les écueils culturels ne sont pas moins importants, liés aux codes et modalités de représentation propres à chaque culture ou ethnie.»

Que faire? Olivier Gainon, fondateur de la maison d’édition électronique CyLibris, explique en décembre 2000: «La première étape est le respect des particularismes au niveau technique. Il faut que le réseau respecte les lettres accentuées, les lettres spécifiques, etc. Je crois très

important que les futurs protocoles permettent une transmission parfaite de ces aspects, ce qui n'est pas forcément simple (dans les futures évolutions de l'HTML ou des protocoles IP, etc.). Donc il faut que chacun puisse se sentir à l'aise avec l'internet et qu'il ne soit pas simplement réservé à des (plus ou moins) anglophones. Il est anormal aujourd'hui que la transmission d'accents puisse poser problème dans les courriers électroniques. La première démarche me semble donc une démarche technique. Si on arrive à faire cela, le reste en découle: la représentation des langues se fera en fonction du nombre de connectés, et il faudra envisager à terme des moteurs de recherche multilingues.»

L'internationalisation

«Vers la communication sur l'internet dans toutes les langues…», tel est le sous-titre de la page d'accueil de Babel, une initiative lancée en 1997 par l'Internet Society (ISOC) et Alis Technologies, société spécialisée dans le traitement automatique des langues. Babel offre un site plurilingue en sept langues (anglais, allemand, espagnol, français, italien, portugais, suédois) qui comprend par exemple des glossaires typographiques et linguistiques et des informations utiles pour développer un site web multilingue, y compris pour l'encodage.

Babel mène la première étude sur la répartition des langues sur le web et publie cette étude en juin 1997 dans les sept langues du site. Le «Palmarès des langues de la Toile» donne les pourcentages de 82,3% pour l'anglais, 4% pour l'allemand, 1,6% pour le japonais, 1,5% pour le français, 1,1% pour l'espagnol, 1,1% pour le suédois et 1% pour l'italien.

Selon la société marketing Global Reach, les 56 millions d'usagers non anglophones recensés en juillet 1998 sont par ordre d'importance les usagers hispanophones (22,4%), les usagers nippons (12,3%), les usagers germanophones (14%) et les usagers francophones (10%).

Randy Hobler, consultant en marketing internet des produits et services de traduction, explique en septembre 1998 dans un entretien par courriel: «Comme l'internet n'a pas de frontières nationales, les usagers s'organisent selon d'autres critères propres au médium. En termes de multilinguisme, vous avez des communautés virtuelles, par exemple ce que j'appelle les "nations de langues", tous ces usagers qu'on peut regrouper selon leur langue maternelle quel que soit leur lieu géographique. Ainsi la nation de la langue espagnole inclut non seulement les internautes d'Espagne et d'Amérique latine, mais aussi tous le Hispanophones vivant aux États-Unis, ou encore ceux qui parlent espagnol au Maroc. L'augmentation de pages web dans d'autres langues que l'anglais n'est pas seulement due au fait qu'il y ait davantage de sites et d'usagers dans des pays non anglophones. Elle est également due au fait que les sociétés et les organisations localisent davantage leurs sites web et au fait qu'on utilise davantage la traduction automatisée pour proposer des sites web à partir ou vers d'autres langues.»

Yoshi Mikami, informaticien à Fujisawa (Japon), écrit (avec ses collègues Kenji Sekine et Nobutoshi Kohara) le premier livre au monde expliquant comment créer des pages web multilingues. «Pour un web multilingue» est publié en japonais par O'Reilly Media en août

1997 avant d’être traduit en anglais, en allemand et en français l’année suivante.

Yoshi Mikami relate en décembre 1998: «Ma langue maternelle est le japonais. Comme j’ai suivi mes études de troisième cycle aux États-Unis et que j’ai travaillé dans l’informatique, je suis devenu bilingue japonais / anglais américain. J’ai toujours été intéressé par différentes langues et cultures, aussi j’ai appris le russe, le français et le chinois dans la foulée. Fin 1995, j’ai créé sur le web la page “The Languages of the World by Computers and the Internet” [Les langues du monde via les ordinateurs et l’internet] et j’ai tenté de proposer — en anglais et en japonais — un bref historique de toutes ces langues, ainsi que les caractéristiques propres à chaque langue et à sa phonétique. Suite à l’expérience acquise, j’ai invité mes deux associés à se joindre à moi pour écrire un livre sur la conception, la création et la présentation de pages web multilingues, livre qui fut publié en août 1997 [en japonais] sous le titre “Pour un web multilingue”, le premier livre au monde sur un tel sujet.»

«Il y a des milliers d’années de cela, en Égypte, en Chine et ailleurs, les gens étaient plus sensibles au fait de communiquer leurs lois et leurs réflexions non seulement dans une langue donnée mais aussi dans plusieurs autres langues. Dans notre monde moderne, chaque pays a adopté plus ou moins une seule langue de communication. À mon avis, l’internet verra l’utilisation plus grande de langues différentes et de pages multilingues — et pas seulement une gravitation autour de l’anglais américain — et un usage plus créatif de la traduction informatique multilingue. 99% des sites web créés au Japon sont en japonais!»

Un consortium industriel international est institué en octobre 1994 pour développer les protocoles communs du web, sous le nom de World Wide Consortium (W3C) et sous l’égide de Tim Berners-Lee, inventeur du web en 1990. Le site web du W3C propose entre autres une page «Internationalization / Localization» avec la liste des protocoles nécessaires à la création d’un site web multilingue (HTML, jeux de caractères, HTTP, URL et autres identificateurs incluant des caractères non ASCII) et des conseils pratiques pour créer un site web multilingue.

Le multilinguisme

Sur les 500 millions d’habitants que compte l’Europe en 1998, 15% sont de langue maternelle anglaise, 28% ne parlent pas l’anglais et 32% utilisent le web en anglais.

Geoffrey Kingscott, directeur de Praetorius, société londonienne de traduction et de services linguistiques, explique en septembre 1998 dans un entretien par courriel: “Les caractéristiques propres au web sont la multiplicité des sites et le bas prix de la diffusion de l’information sur ces sites. Ceci favorisera donc le multilinguisme au fur et à mesure du développement du web. Comme le web s’est d’abord développé aux États-Unis, il est encore principalement en anglais, mais ce n’est qu’un phénomène temporaire. Pour expliquer ceci plus en détail, je dirais que, lorsque nous étions dépendants de l’imprimé ou de l’audiovisuel (films, télévision, radio, vidéos, cassettes), l’information ou le divertissement que nous attendions dépendait d’agents (éditeurs, stations de télévision ou de radio, producteurs de cassettes ou de vidéos) qui devaient

subsister commercialement et, dans le cas de la radio-télédiffusion du service public, avec de sévères contraintes budgétaires. Le nombre de clients était primordial, et déterminait la nécessité de langues autres que l'omniprésent anglais. Ces contraintes disparaissent avec le web.»

Alain Bron, consultant en systèmes d'information et écrivain basé à Paris, explique en janvier 1999: «Il y aura encore pendant longtemps l'usage de langues différentes et tant mieux pour le droit à la différence. Le risque est bien entendu l'envahissement d'une langue au détriment des autres, donc l'aplanissement culturel. Je pense que des services en ligne vont petit à petit se créer pour pallier cette difficulté. Tout d'abord, des traducteurs pourront traduire et commenter des textes à la demande, et surtout les sites de grande fréquentation vont investir dans des versions en langues différentes, comme le fait l'industrie audiovisuelle.»

Marcel Grangier, responsable de la section française des services linguistiques centraux de l'Administration fédérale suisse, écrit en janvier 1999: "Le multilinguisme sur l'internet peut être considéré comme une fatalité heureuse et surtout irréversible. C'est dans cette optique qu'il convient de creuser la tombe des rabat-joie dont le seul discours est de se plaindre d'une suprématie de l'anglais. Cette suprématie n'est pas un mal en soi, dans la mesure où elle résulte de réalités essentiellement statistiques (plus de PC par habitant, plus de locuteurs de cette langue, etc.). La riposte n'est pas de "lutter contre l'anglais" et encore moins de s'en tenir à des jérémiades, mais de multiplier les sites en d'autres langues. Notons qu'en qualité de service de traduction, nous préconisons également le multilinguisme des sites eux-mêmes. La multiplication des langues présentes sur l'internet est inévitable, et ne peut que bénéficier aux échanges multiculturels. Pour que ces échanges prennent place dans un environnement optimal, il convient encore de développer les outils qui amélioreront la compatibilité. La gestion complète des diacritiques ne constitue qu'un exemple de ce qui peut encore être entrepris.»

Bruno Didier, webmestre de la bibliothèque de l'Institut Pasteur à Paris, écrit en août 1999: «L'internet n'est une propriété ni nationale, ni linguistique. C'est un vecteur de culture, et le premier support de la culture, c'est la langue. Plus il y a de langues représentées dans leur diversité, plus il y aura de cultures sur l'internet. Je ne pense pas qu'il faille justement céder à la tentation systématique de traduire ses pages dans une langue plus ou moins universelle. Les échanges culturels passent par la volonté de se mettre à la portée de celui vers qui on souhaite aller. Et cet effort passe par l'appréhension de sa langue. Bien entendu c'est très utopique comme propos. Concrètement, lorsque je fais de la veille, je peste dès que je rencontre des sites norvégiens ou brésiliens sans un minimum d'anglais.»

Steven Krauwer, coordinateur d'ELSNET (European Network of Excellence in Human Language Technologies — Réseau européen d'excellence dans les technologies du langage humain), écrit en septembre 1998: «En tant que citoyen européen, je pense que le multilinguisme sur le web est absolument essentiel. À mon avis, ce n'est pas une situation saine à long terme que seuls ceux qui ont une bonne maîtrise de l'anglais puissent pleinement exploiter les bénéfices du web. En tant que chercheur (spécialisé dans la traduction automatique), je vois le multilinguisme comme un défi majeur: pouvoir garantir que l'information sur le web soit accessible à tous, indépendamment des différences de langue.»

Il ajoute en août 1999: «Je suis de plus en plus convaincu que nous devons veiller à ne pas aborder le problème du multilinguisme en l'isolant du reste. Je reviens de France, où j'ai passé de très bonnes vacances d'été. Même si ma connaissance du français est sommaire (c'est le moins que l'on puisse dire), il est surprenant de voir que je peux malgré tout communiquer sans problème en combinant ce français sommaire avec des gestes, des expressions du visage, des indices visuels, des schémas, etc. Je pense que le web (contrairement au système vieillot du courrier électronique textuel) peut permettre de combiner avec succès la transmission des informations par différents canaux (ou différents moyens), même si ce processus n'est que partiellement satisfaisant pour chacun des canaux pris isolément.»

Selon lui, comment promouvoir le multilinguisme sur le web? «Pour les créateurs de sites web, une meilleure formation pour exploiter les combinaisons de modalités possibles afin d'améliorer la communication par-delà la barrière des langues (et pas seulement par un vernis superficiel). Pour les usagers, des logiciels de traduction comme AltaVista Translation [plus connu sous le nom de Babel Fish], dont la qualité n'est pas frappante, mais qui a le mérite d'exister. Pour les navigateurs web, des logiciels de traduction intégrée, particulièrement pour les langues non dominantes, et des dictionnaires intégrés plus rapides.»

La localisation des sites web

Le moteur de recherche Yahoo! est le premier à proposer fin 1997 une interface en sept langues (anglais, allemand, coréen, français, japonais, norvégien, suédois), très appréciée de ses usagers non anglophones, dont le nombre ne cesse de croître.

Selon Brian King, directeur du WorldWide Language Institute (Institut des langues du monde entier), interviewé en septembre 1998: «Bien qu'un web multilingue soit souhaitable sur le plan moral et éthique, un tel idéal ne suffit pas pour en faire une réalité transcendant les limites actuelles. De même que l'utilisateur non anglophone peut maintenant avoir accès aux technologies dans sa propre langue, l'impact du commerce électronique peut devenir une force majeure qui fasse du multilinguisme la voie la plus naturelle vers le cyberespace. Les vendeurs de produits et services dans le marché virtuel mondial que devient l'internet doivent être préparés à faire face à un monde virtuel qui soit aussi multilingue que le monde physique. S'ils veulent réussir, ils doivent s'assurer qu'ils parlent bien la langue de leurs clients!»

C'est ce que fait Bill Dunlap, qui fonde d'abord en 1985 Euro-Marketing Associates, une société de conseil en marketing basée à Paris et San Francisco. Il restructure cette société en 1995 pour créer Global Reach, un service de conseil en ligne dont le but est de promouvoir les sites web des entreprises américaines en Europe, afin d'attirer plus de visiteurs, et donc d'augmenter les ventes. Le service de conseil comprend la traduction des sites dans plusieurs langues, la promotion active des sites traduits et l'accroissement de la fréquentation locale au moyen de bandeaux publicitaires ciblés.

Bill Dunlap écrit en décembre 1998 lors d’un entretien par courriel: «Il y a très peu d’Américains vivant aux États-Unis qui sont intéressés de communiquer dans plusieurs langues. Pour la plupart, ils pensent encore que le monde entier parle anglais. Par contre, ici en Europe (j’écris de Paris), les pays sont petits, si bien que, depuis des siècles, une perspective internationale est nécessaire. (…) Depuis 1981, début de mon activité professionnelle, j’ai été impliqué dans la venue de sociétés américaines en Europe. Ceci est pour beaucoup un problème de langue, puisque leurs informations commerciales doivent être disponibles dans les langues européennes pour être prises en compte en Europe. Comme le web est devenu populaire en 1995, j’ai donné à ces activités une dimension “en ligne”, et j’en suis venu à promouvoir le cyber-commerce européen auprès de mes compatriotes américains.»

Peter Raggett, directeur de la bibliothèque de l’OCDE (Organisation de coopération et de développement économiques) à Paris, écrit en août 1999: «Je pense qu’il appartient aux organisations et sociétés européennes d’offrir des sites web si possible en trois ou quatre langues. À l’heure de la mondialisation et du commerce électronique, les sociétés ont un marché potentiel sur plusieurs pays à la fois. Permettre aux usagers francophones, germanophones ou japonais de consulter un site web aussi facilement que les usagers anglophones donnera une plus grande compétitivité à ces sociétés.»

La collaboration en ligne

Lors d’un entretien mené en décembre 1997 pour le quotidien suisse La Tribune de Genève, Pierre Ruetschi, journaliste, pose la question suivante à Tim Berners-Lee, inventeur du web en 1990: «Sept ans plus tard, êtes-vous satisfait de la façon dont le web a évolué?». Tim Berners-Lee répond en substance que, s’il est heureux de la richesse et de la variété de l’information disponible, il aimerait «que le web soit plus interactif, que les gens puissent créer de l’information ensemble», et pas seulement consommer celle qui leur est proposée. Le web doit devenir «un média de collaboration, un monde de connaissance que nous partageons».

Ingénieure d’études à l’INaLF (Institut national de la langue française), Christiane Jadelot raconte en juillet 1998: «J’ai commencé à utiliser vraiment l’internet en 1994, je crois, avec un logiciel qui s’appelait Mosaic. J’ai alors découvert un outil précieux pour progresser dans mes connaissances en informatique, linguistique, littérature… Tous les domaines sont couverts. Il y a le pire et le meilleur, mais en consommateur averti, il faut faire le tri de ce que l’on trouve. J’ai surtout apprécié les logiciels de courrier, de transfert de fichiers, de connexion à distance. J’avais à cette époque des problèmes avec un logiciel qui s’appelait Paradox et des polices de caractères inadaptées à ce que je voulais faire. J’ai tenté ma chance et posé la question dans un groupe de News approprié. J’ai reçu des réponses du monde entier, comme si chacun était soucieux de trouver une solution à mon problème!»

Basé dans la Silicon Valley en Californie, Murray Suid écrit des livres pédagogiques, des livres pour enfants, des oeuvres multimédias et des scénarios. Il raconte en juillet 1998 dans un entretien par courriel: «L’internet est devenu mon principal instrument de recherche, et il a

largement — mais pas complètement — remplacé la bibliothèque traditionnelle et la communication de personne à personne pour une recherche précise. À l'heure actuelle, au lieu de téléphoner ou d'aller interviewer les gens sur rendez-vous, je le fais par courrier électronique. Du fait de la rapidité inhérente à la messagerie électronique, j'ai pu collaborer à distance avec des gens, particulièrement pour des scénarios. J'ai par exemple travaillé avec deux producteurs allemands. Cette correspondance est également facile à conserver et à organiser, et je peux donc aisément accéder à l'information échangée de cette façon. De plus, le fait d'utiliser le courrier électronique permet aussi de garder une trace des idées et des références documentaires. Ce type de courrier fonctionnant bien mieux que le courrier classique, l'internet m'a permis de beaucoup augmenter ma correspondance.»

«De même, le rayon géographique de mes correspondants s'est beaucoup étendu, surtout vers l'Europe. Auparavant, j'écrivais rarement à des correspondants situés hors des États-Unis. C'est également beaucoup plus facile, et je prends nettement plus de temps qu'avant pour aider d'autres écrivains dans une sorte de groupe de travail virtuel. Ce n'est pas seulement une attitude altruiste. J'apprends beaucoup de ces échanges qui, avant l'internet, me demandaient beaucoup plus d'efforts. Je n'aurais jamais pu préparer mon dernier livre sans utiliser le courrier électronique parce que cela m'aurait coûté trop de temps et d'argent pour localiser les experts.»

Créé en 1996 par Robert Ware, informaticien américain, OneLook Dictionaries est un moteur de recherche puisant dans 2 millions de mots provenant de 425 dictionnaires (en août 1998) dans divers domaines (dictionnaires généralistes, affaires, informatique / internet, technologies, sciences, médecine, religion, sports, argot) et dans plusieurs langues (anglais, français, allemand, italien, espagnol).

Robert Ware raconte en septembre 1998: "À titre personnel, je suis presque uniquement en contact avec des personnes qui ne parlent qu'une langue et qui ne sont pas très motivées pour développer leurs aptitudes linguistiques. Être en contact avec le monde entier change cette approche des choses. Et la change en mieux! J'ai été long à inclure des dictionnaires non anglophones (en partie parce que je suis monolingue). Mais vous en trouverez maintenant quelques-uns.»

Robert Ware raconte aussi dans le même entretien: «Un fait intéressant s'est produit dans le passé qui a été très instructif pour moi. En 1994, je travaillais pour un établissement scolaire et je tentais d'installer un logiciel sur un modèle d'ordinateur particulier. J'ai trouvé une personne qui avait les mêmes difficultés, et nous avons commencé à échanger des courriers électroniques. Soudain, cela m'a frappé… Le logiciel avait été écrit à 40 km de là [dans l'État du Colorado aux États-Unis], mais c'était une personne située de l'autre côté de la planète qui m'aidait. Les distances et les considérations géographiques n'importaient plus! En effet c'est épatant, mais à quoi cela nous mène-t-il? Je ne peux communiquer qu'en anglais mais, heureusement, mon correspondant pouvait utiliser aussi bien l'anglais que l'allemand qui était sa langue maternelle. L'internet a supprimé une barrière, celle de la distance, mais il subsiste la barrière de la langue, bien réelle.»

«Il semble que l'internet propulse simultanément les gens dans deux directions différentes. L'internet, anglophone à l'origine, relie les gens dans le monde entier. Par là même il favorise une langue commune pour communiquer. Mais il crée aussi des contacts entre des personnes de

langues différentes et suscite ainsi un intérêt plus grand pour le multilinguisme. Si une langue commune est appréciable, elle ne remplace en aucun cas la nécessité du multilinguisme. L'internet favorise ainsi à la fois une langue commune et le multilinguisme, et ceci est un facteur qui aide à trouver des solutions. L'intérêt croissant pour les langues et le besoin que nous avons de ces langues favorisent de par le monde la création de cours de langues et d'outils linguistiques, et l'internet fournit la possibilité de les rendre disponibles rapidement et à bon marché.»

Tout aussi important, le web doit être accessible à tous, y compris — et surtout — aux personnes en situation de handicap. C'est la tâche que s'est donnée l'association HandicapZéro, créée en 1987 à partir du constat suivant: l'information visuelle est omniprésente, mais les personnes aveugles et malvoyantes n'y ont pas accès.

Mis en ligne en septembre 2000, le site web de l'association devient rapidement le site adapté le plus visité d'Europe, avec 10.000 requêtes mensuelles. HandicapZéro entend ainsi démontrer «que, sous réserve du respect de certaines règles élémentaires, l'internet peut devenir enfin un espace de liberté pour tous.»

Suite à cette première expérience réussie, HandicapZéro lance en février 2003 un portail généraliste offrant en accès libre l'information nationale et internationale en temps réel (en partenariat avec l'Agence France-Presse), l'actualité sportive (avec le journal L'Équipe), les programmes de télévision (avec le magazine Télérama), la météo (avec Météo France), un moteur de recherche (avec Google) et toute une gamme de services (santé, emploi, consommation, loisirs et téléphonie), avec 200.000 requêtes mensuelles en 2006.

Les usagers aveugles peuvent accéder au site au moyen d'une plage braille ou d'une synthèse vocale. Les usagers malvoyants peuvent paramétrer la taille et la police des caractères ainsi que la couleur du fond d'écran pour une navigation confortable, en créant puis en modifiant leur profil selon leur potentiel visuel. Ce profil peut être utilisé pour lire tout texte situé sur le web, en faisant un copier-coller dans la fenêtre prévue à cet effet. Les usagers voyants peuvent correspondre en braille avec les usagers aveugles par le biais du site, HandicapZéro assurant gratuitement la transcription et l'impression braille des courriers (de 4.000 caractères maximum) ainsi que leur expédition par voie postale dans les pays de l'Union européenne.

Les textes électroniques

Les premiers textes électroniques disponibles sur le web sont des e-zines (zines électroniques). John Labovitz, créateur de l'E-Zine-List en 1993, explique sur son site: «Pour ceux d'entre vous qui ne connaissent pas le monde du zine, "zine" est l'abrégé de "fanzine" ou "magazine" selon votre point de vue. Les zines sont en général l'oeuvre d'une personne ou d'un petit groupe, souvent rédigée pour le plaisir ou pour des raisons personnelles, et sont le plus souvent irrévérencieux, bizarres et/ou ésotériques. Les zines ne sont pas des publications grand public. Le plus souvent ils ne contiennent pas de publicité (sauf parfois des publicités pour d'autres

zines), ils ne sont pas dirigés vers une audience de masse et ils ne visent pas un profit commercial. Un “e-zine” est un zine qui est distribué en partie ou uniquement sur des réseaux électroniques tels que l’internet.»

Les premiers e-zines sont diffusés dans les Etext Archives, créées en 1992 par Paul Southworth et hébergées sur le site web de l’Université du Michigan (États-Unis).

Les Etext Archives sont «un lieu d’accueil pour les textes électroniques de tout genre, du sacré au profane, et du politique au personnel», sans juger de leur contenu. Cinq ans plus tard, elles comportent six sections: (1) une section «E- zines», qui regroupe des textes électroniques périodiques qui vont du professionnel au personnel; (2) une section «Politics», qui regroupe des e-zines politiques, ainsi que des essais et des pages web de groupes politiques; (3) une section «Fiction», qui regroupe des publications d’auteurs amateurs; (4) une section «Religion», qui regroupe des textes religieux grand public ou non; (5) une section «Poetry», qui est un mélange éclectique de poésie surtout amateur; (6) une section «Quartz», qui comprend les archives auparavant hébergées à quartz.rutgers.edu.

Comme indiqué sur le site en 1998: «Le web venait de débuter, le gopher [menu textuel à plusieurs niveaux] était la nouvelle technologie de pointe et le FTP [protocole de transfert de fichiers] était encore le protocole standard d’extraction de l’information pour la grande majorité des utilisateurs. La manière dont le projet a débuté a incité de nombreuses personnes à l’associer avec l’Université du Michigan, bien qu’il n’y ait aucune relation officielle avec l’université et bien que le projet soit entièrement le fait du travail des volontaires et de dons personnels. Le matériel informatique est la propriété exclusive des responsables du projet.»

«Le projet a été lancé en réponse à l’absence d’archivage organisé de documents politiques, de périodiques et de discussions diffusées par le biais de Usenet sur des newsgroups. Peu de temps après, les e-zines ont débuté leur prolifération rapide sur l’internet, et il était clair que ces publications souffraient de la même absence de collecte coordonnée et de préservation, sans parler du fait que la frontière était floue entre les e-zines (qui à l’époque était surtout liés au hacking, au phreaking et à l’anarchisme internet) et les documents politiques présents sur l’internet, si bien que la plupart des e-zines étaient en phase avec l’objectif original des Etext Archives. Une chose en amenant une autre, des e-zines de toutes sortes — dont de nombreux titres sur divers sujets culturels non liés à la politique — ont fini par envahir nos archives en volume significatif.»

Un autre répertoire est la E-Zine-List, créée en été 1993 par John Labovitz pour recenser les e-zines accessibles par FTP, gopher et courriel ou bien disponibles sur le web, et ceci dans le monde entier. La liste est actualisée une fois par mois. Dans l’historique présent sur le site en 1998, John Labovitz relate qu’à l’origine son intention est de faire connaître Crash, un zine imprimé dont il souhaite proposer une version électronique. À la recherche de répertoires, il ne trouve que le groupe de discussion alt.zines et des archives comme The Well et les Etext Archives. Lui vient alors l’idée d’un répertoire organisé, qu’il débute avec douze titres classés manuellement dans un traitement de texte. Puis il écrit sa propre base de données.

En cinq ans, de 1993 à 1998, les e-zines passent de quelques dizaines à plusieurs centaines, et la signification même d'e-zine s'élargit pour recouvrir tout type de publication publiée par voie électronique, même si, selon John Labovitz, «il subsiste toujours un groupe original et indépendant désormais minoritaire qui continue de publier suivant son coeur ou de repousser les frontières de ce que nous appelons un zine.» La E-Zine-List recense 3.045 titres en novembre 1998. John Labovitz poursuit sa liste jusqu'en novembre 2001 avant de passer le relais à d'autres.

Les textes électroniques sont aussi des livres du domaine public saisis page après page sur un clavier d'ordinateur et diffusés gratuitement sur le web. Ces livres sont recensés dès 1993 sur The Online Books Page (La page des livres en ligne) par John Mark Ockerbloom, doctorant à l'Université Carnegie Mellon (États-Unis), afin d'offrir au lecteur un point d'accès commun.

John Mark Ockerbloom relate en septembre 1998 dans un entretien par courriel: «J'étais webmestre ici pour la section informatique de la CMU [Carnegie Mellon University], et j'ai débuté notre propre site en 1993. Il comprenait des pages avec des liens vers des ressources disponibles localement, et à l'origine The Online Books Page était l'une de ces pages, avec des liens vers des livres mis en ligne par des collègues de notre département (par exemple Robert Stockton, qui a créé les versions web de certains textes du Projet Gutenberg). Ensuite les gens ont commencé à demander des liens vers des livres disponibles sur d'autres sites. J'ai remarqué que de nombreux sites (et pas seulement le Projet Gutenberg ou Wiretap) proposaient des livres en ligne, et qu'il serait utile d'en avoir une liste complète qui permette de télécharger et lire des livres où qu'ils soient sur l'internet. C'est ainsi que mon index a débuté.»

«J'ai quitté mes fonctions de webmestre en 1996, mais j'ai gardé la gestion de The Online Books Page, parce que je m'étais entre temps passionné pour l'énorme potentiel que possède l'internet de rendre la littérature accessible au plus grand nombre. Maintenant il y a tant de livres mis en ligne que j'ai du mal à rester à jour. Je pense pourtant poursuivre cette activité d'une manière ou d'une autre. Je suis très intéressé par le développement de l'internet en tant que médium de communication de masse dans les prochaines années. J'aimerais également rester impliqué dans la mise à disposition gratuite de livres sur l'internet, que ceci fasse partie intégrante de mon activité professionnelle, ou que ceci soit une activité bénévole menée sur mon temps libre.»

En 1998, un index de 7.000 livres en ligne est disponible par auteur, titre et sujet. On trouve aussi une liste de répertoires de textes en ligne, tout comme une liste de répertoires de publications périodiques (magazines, journaux, revues grand public, revues scientifiques).

Fin 1998, John Mark Ockerbloom obtient son doctorat en informatique. En 1999, il rejoint l'Université de Pennsylvanie pour contribuer au développement de la bibliothèque numérique de l'université. À la même époque, il y transfère The Online Books Page — qui répertorie déjà 12.000 titres — tout en gardant la même présentation, très sobre, et tout en poursuivant le travail d'inventaire dans le même esprit. The Online Books Page recense 20.000 titres en 2003 (dont 4.000 titres publiés par des femmes), 25.000 titres en 2006, 30.000 titres en 2007 (dont 7.000 titres du Projet Gutenberg), 35.000 titres en 2010 et 2 millions de titres en 2015.

Les livres numériques

Bien avant que les grandes bibliothèques mondiales débutent leurs propres bibliothèques numériques, on voit l'apparition de collections numériques dans diverses langues librement disponibles sur le web.

La plupart s'inspirent du Projet Gutenberg, créé en 1971 par Michael Hart, étudiant à l'Université de l'Illinois (États-Unis), pour mettre gratuitement à la disposition de tous, par voie électronique, le plus grand nombre possible d'oeuvres du domaine public, littéraires ou non. Au lieu d'un ensemble de pages reliées, le livre devient un texte électronique que l'on peut dérouler en continu, avec des lettres capitales pour les termes en italique, en gras et soulignés de la version imprimée.

Michael Hart raconte en août 1998 dans un entretien par courriel: «Nous considérons le texte électronique comme un nouveau médium, sans véritable relation avec le papier. Le seul point commun est que nous diffusons les mêmes oeuvres, mais je ne vois pas comment le papier peut concurrencer le texte électronique une fois que les gens y sont habitués, particulièrement dans les écoles.»

Après la saisie de la Constitution des États-Unis et de textes connexes sur le clavier de son terminal, Michael Hart aborde la Bible, avec un fichier par chapitre, puis les oeuvres complètes de Shakespeare, avec un fichier par oeuvre. Des volontaires se joignent à lui, et le travail se poursuit page après page dans les années 1980 et 1990. Les premiers livres disponibles en français sont six romans de Stendhal et deux romans de Jules Verne, tous mis en ligne début 1997.

Le Projet Gutenberg propose des livres en 25 langues début 2004, en 42 langues (y compris le sanskrit et les langues mayas) en juillet 2005 et en 60 langues en novembre 2010. Les dix langues principales sont l'anglais (avec 28.666 livres le 6 novembre 2010), le français (1.672 livres), l'allemand (715 livres), le finnois (542 livres), le néerlandais (498 livres), le portugais (474 livres), le chinois (405 livres), l'espagnol (297 livres), l'italien (253 livres) et le grec (107 livres). Ces langues sont suivies du latin, de l'espéranto, du suédois et du tagalog.

Cette croissance rapide est due à l'activité de Distributed Proofreaders, un site lancé en octobre 2000 par Charles Franks pour permettre la numérisation des livres par OCR (reconnaissance optique de caractères) puis leur relecture partagée entre de nombreux volontaires. Le nombre total de livres produits par Distributed Proofreaders est de 3.000 livres en février 2004, 5.000 livres en octobre 2004, 7.000 livres en mai 2005, 8.000 livres en février 2006, 10.000 livres en mars 2007 et 30.000 livres en juillet 2015, avec plusieurs milliers de volontaires de par monde et une production de cinq livres par jour.

Michael Hart précise souvent dans ses écrits que, si Gutenberg a permis à chacun d'avoir ses propres livres — jusque-là réservés à une élite — pour un coût relativement modique, le Projet Gutenberg permet à chacun d'avoir une bibliothèque complète gratuite — jusque-là réservée à

une collectivité — sur un support qu'on peut glisser dans sa poche (ou porter en pendentif autour du cou), par exemple une clé USB. En septembre 2015, le Projet Gutenberg propose 50.000 livres numériques dans plusieurs langues, tous relus et corrigés, téléchargeables gratuitement dans plusieurs formats et indéfiniment reproductibles.

La première collection numérique nordique (scandinave) est le Projekt Runeberg, affilié au Projet Gutenberg. Le Projekt Runeberg est créé en décembre 1992 par Lysator, un club informatique d'étudiants, en collaboration avec la bibliothèque de l'Université de Linkoping (Suède), pour produire et organiser des versions électroniques gratuites de la littérature classique nordique. 200 oeuvres sont disponibles en 1998, avec une liste de 6.000 auteurs nordiques en tant qu'outil de développement des collections.

La première collection numérique allemande est le Projekt Gutenberg-DE, lui aussi affilié au Projet Gutenberg. Le Projekt Gutenberg-DE est créé en 1994 dans le même esprit pour offrir des versions électroniques gratuites de la littérature classique allemande. Plusieurs dizaines d'oeuvres sont disponibles en ligne en 1998, avec une page web pour les oeuvres courtes et plusieurs pages — une par chapitre — pour les oeuvres plus longues. Le site propose aussi une liste alphabétique d'auteurs et de titres, ainsi qu'une courte biographie et une bibliographie pour chaque auteur.

La première collection numérique française est ABU-La bibliothèque universelle, créée en avril 1993 à l'initiative de l'Association des bibliophiles universels (ABU) et hébergée par le Centre d'études et de recherche informatique (CEDRIC) du Conservatoire des arts et métiers (CNAM) de Paris. Le terme ABU est inspiré d'Aboulafia, petit ordinateur présent dans «Le pendule de Foucault», un roman de Umberto Eco dont l'intrigue se situe justement au CNAM. Quant au nom de l'association, «au départ, il s'agissait de biblioFiles universels, et non de biblioPHiles, mais la préfecture de Paris n'a pas semblé saisir tout le sel de ce néologisme», explique l'ABU sur son site. Le site permet l'accès libre au texte intégral d'oeuvres du domaine public soit, en chiffres, 223 oeuvres de 76 auteurs en 1998 et 288 oeuvres de 101 auteurs en 2002.

La première collection numérique suisse est Athena, créée en 1994 par Pierre Perroud, professeur au collège Voltaire de Genève, et hébergée sur le site de l'Université de Genève. En décembre 1997, le site bilingue français-anglais propose à la fois 200 oeuvres numérisées par Athena et des liens vers des oeuvres en accès libre sur le web dans des domaines aussi variés que la philosophie, les sciences, la période classique, la littérature, l'histoire ou l'économie, avec 3.500 titres dans plusieurs langues en décembre 1997 et 8.000 titres un an plus tard. Athena se veut multilingue, avec une section dédiée aux auteurs de langue française et une autre section dédiée aux auteurs suisses puisque Genève est la capitale de la Suisse francophone. Athena propose aussi un répertoire mondial de ressources littéraires en ligne.

Pierre Perroud insiste en février 1997 dans un article de la revue suisse Informatique – Informations sur la complémentarité du texte électronique et du livre imprimé. Selon lui, le texte électronique encourage la lecture et la diffusion de la culture, et il est donc un bon complément du livre imprimé, qui reste irremplaçable lorsqu'il s'agit de lire. Le livre imprimé reste «un compagnon mystérieusement sacré vers lequel convergent de profonds symboles: on le serre dans la main, on le porte contre soi, on le regarde avec admiration; sa petitesse nous rassure

autant que son contenu nous impressionne; sa fragilité renferme une densité qui nous fascine; comme l'homme il craint l'eau et le feu, mais il a le pouvoir de mettre la pensée de celui-là à l'abri du Temps.»

La première collection numérique italienne est le Progetto Manuzio, créé en 1995 par Liber Liber, une association promouvant l'expression créatrice artistique et intellectuelle, et utilisant les technologies informatiques pour lier les humanités aux sciences. Le nom du projet est inspiré du nom d'un éditeur vénitien du 16e siècle connu pour avoir amélioré les techniques d'impression inventées par Gutenberg. Comme indiqué sur le site, le Progetto Manuzio veut «rendre une noble idée réelle: l'idée de mettre la culture à la disposition de tous. Comment? En mettant des livres, thèses, articles, oeuvres de fiction ou tout autre document numérisable à la disposition du monde entier, à n'importe quel moment, et gratuitement. Par le biais d'un modem et de disquettes (dans ce cas, en ajoutant le coût d'une disquette vierge et des frais postaux), il est déjà possible d'obtenir des centaines de livres. Et le Progetto Manuzio n'a besoin que de quelques personnes pour mettre un chef-d'oeuvre tel que "La Divine Comédie" de Dante à la disposition de millions de gens.»

La presse

Au début des années 1990, les premières éditions électroniques de journaux transitent d'abord par des services commerciaux comme America OnLine (AOL) ou CompuServe. Puis les grands titres de la presse imprimée lancent leurs propres sites web dans nombre de pays.

Mis en ligne en février 1995, le site web du mensuel Le Monde diplomatique est le premier site d'un périodique imprimé français, suivi du quotidien Libération fin 1995 et du quotidien Le Monde en 1996. Avec un site web lancé la même année, L'Humanité est le premier quotidien français à proposer la version intégrale du journal en accès libre. La presse régionale est tout aussi présente. Le quotidien Les Dernières nouvelles d'Alsace lance son site web en septembre 1995 et le quotidien Ouest-France lance son site web en juillet 1996, parmi tant d'autres.

Au Royaume-Uni, le Times et le Sunday Times font web commun sur le site Times Online, avec possibilité de créer une édition personnalisée. Aux États-Unis, la version en ligne du Wall Street Journal est payante, avec 100.000 abonnés en 1998. Celle du New York Times est disponible sur abonnement gratuit. Le Washington Post propose l'actualité du jour en accès libre ainsi que de nombreux articles archivés, le tout avec images, sons et vidéos. Pathfinder (qui devient ensuite Time) est le site web du groupe Time-Warner, éditeur de Time Magazine, Sports lllustrated, Fortune, People, Southern Living, Money, Sunset, etc. On peut lire tous les articles de ces magazines et les rechercher par date ou par sujet. Lancé en 1992 en Californie, Wired, premier magazine imprimé entièrement consacré à la culture cyber, est bien évidemment présent sur le web.

Autre apport de taille, l’internet permet de lire en ligne des titres difficiles ou impossibles à trouver en kiosque. Un article du quotidien Le Monde du 23 mars 1998 donne l’exemple du quotidien algérien El Watan, en ligne depuis octobre 1997. Selon Redha Belkhat, rédacteur en chef du journal, l’enjeu est capital: «Pour la diaspora algérienne, trouver dans un kiosque à Londres, New York ou Ottawa, un numéro d’El Watan daté de moins d’une semaine relève de l’exploit. Maintenant, le journal tombe ici à 6 heures du matin, et à midi il est sur internet.»

Dans un essai publié en décembre 1997 sur le site de l’AJR/NewsLink, le journaliste Eric K. Meyer explique: «Plus de 3.600 journaux ont maintenant une présence sur l’internet. 43% des journaux en ligne sont basés hors des États-Unis, alors qu’ils n’étaient que 29% il y a un an. Une croissance rapide — surtout au Canada, au Royaume-Uni, en Norvège, au Brésil et en Allemagne — a fait grimper à 1.563 titres le nombre de journaux en ligne qui ne sont pas publiés aux États-Unis. Le nombre de journaux américains en ligne est lui aussi en hausse et il est passé de 745 journaux il y a un an à 1.290 journaux il y a six mois puis à 2.059 journaux aujourd’hui. Hors des États-Unis, les pays leaders sont le Royaume-Uni avec 294 journaux en ligne et le Canada avec 230 journaux en ligne. Au Canada, chaque province ou territoire dispose d’au moins un journal en ligne. La province la plus représentée est l’Ontario avec 91 journaux, suivie par l’Alberta avec 44 journaux puis par la Colombie-Britannique avec 43 journaux. Ailleurs en Amérique, le Mexique dispose de 51 journaux en ligne, l’Amérique centrale de 23 journaux en ligne et les Caraïbes de 36 journaux en ligne. L’Europe est le deuxième continent le plus actif, avec 728 journaux en ligne. Le pays leader est le Royaume-Uni, suivi de la Norvège avec 53 journaux en ligne et l’Allemagne avec 43 journaux en ligne. L’Asie dispose de 223 journaux en ligne avec l’Inde comme pays leader. L’Amérique du Sud dispose de 161 journaux en ligne avec la Bolivie comme pays leader. L’Afrique dispose de 53 journaux en ligne avec l’Afrique du Sud comme pays leader. L’Australie et les îles voisines disposent de 64 journaux en ligne.»

Le web permet aussi d’approfondir tous ces articles, avec un accès instantané à une foule d’informations telles que cartes géographiques, notes biographiques, textes officiels, données politiques et économiques, photos, bandes sonores et vidéos.

Les lettres d’information électroniques sans équivalent imprimé sont de plus en plus nombreuses, par exemple la lettre d’information de Jacques Gauchey, journaliste et écrivain basé dans la Silicon Valley en Californie. Après avoir travaillé entre 1985 et 1992 pour le quotidien La Tribune, Jacques Gauchey crée en 1993 la société G. a Communications, qui facilite les relations entre les sociétés de la côte ouest des États-Unis et leurs partenaires européens.

Jacques Gauchey raconte en juillet 1999 dans un entretien par courriel: «J’ai publié quelques numéros d’une lettre d’information en anglais gratuite il y a quatre ans sur l’internet. Une dizaine de lecteurs par numéro jusqu’au jour (en janvier 1996) où l’édition électronique de Wired créa un lien. En une semaine j’ai eu une centaine de courriers électroniques — y compris de lecteurs français de mon livre “La vallée du risque: Silicon Valley” [publié par Plon en 1990], contents de me retrouver.»

La précarisation de l’emploi

Tout contenu est désormais systématiquement numérisé pour permettre son transfert par voie électronique, ce qui accélère le processus matériel de production. Dans la presse, alors qu’auparavant le personnel de production devait dactylographier les textes du personnel de rédaction, les journalistes envoient directement leurs textes pour mise en page. Dans l’édition, le rédacteur, le concepteur artistique et l’infographiste travaillent souvent simultanément sur le même ouvrage.

Si certains secteurs créent de nouveaux emplois, par exemple l’industrie audiovisuelle, d’autres secteurs sont soumis à des restructurations ou même à des licenciements en masse. Ces problèmes sont suffisamment préoccupants pour être débattus lors du premier colloque sur la convergence multimédia organisé en janvier 1997 par le Bureau international du travail (BIT) à Genève. Ce colloque rassemble des employeurs, des syndicalistes et des représentants des gouvernements venus du monde entier.

Selon Wilfred Kiboro, président de Nation Printers and Publishers (Imprimeurs et éditeurs de la nation) au Kenya, «le coût de la technologie de l’information doit être ramené à un niveau abordable. Je rêve du jour où les villageois africains pourront accéder à l’internet depuis leur village, aujourd’hui privé d’eau et d’électricité.» Les problèmes économiques sont doublés de problèmes culturels. Paradoxalement, les informations concernant l’Afrique à destination des Africains ne viennent pas du continent lui-même, mais elles sont diffusées par de grands médias occidentaux qui transmettent leur propre vision de l’Afrique, sans réelle perception de sa situation économique et sociale.

Un problème tout aussi préoccupant est celui de la pression constante exercée sur les journalistes des salles de rédaction, comme expliqué par Bernie Lunzer, secrétaire-trésorier de la Newspaper Guild (Guilde des journaux) aux États-Unis. Le produit de leur travail n’est plus utilisé seulement une fois en fin de journée, mais plusieurs fois par jour. Ces tensions à répétition sont encore aggravées par une journée de travail à l’écran pendant huit à dix heures d’affilée. Le rythme de travail et l’utilisation intensive de l’ordinateur entraînent des problèmes de santé et de sécurité au travail, avec des journalistes qui craquent à l’âge de 35 ou 40 ans après quelques années à ce régime.

Cette pression est dénoncée aussi par Carlos Alberto de Almeida, président de la Fédération nationale des journalistes (FENAJ) du Brésil. Selon lui, les nouvelles technologies devraient donner la possibilité de rationaliser le travail et d’en réduire la durée pour favoriser l’enrichissement intellectuel et les loisirs. Mais en pratique les journalistes voient leur nombre d’heures de travail augmenter, avec des journées de travail de dix à douze heures. Les heures supplémentaires ne sont pas payées, comme ne sont pas payées non plus les heures de travail non prévues le week-end alors que les journalistes sont censés être en période de repos.

Etienne Reichel, directeur suppléant de l'association suisse Viscom (Visual Communication — Communication visuelle), explique que l'utilisation de l'internet entraîne la suppression de certaines phases de production et réduit le nombre d'emplois. Le travail de vingt typographes est désormais assuré par six travailleurs qualifiés, alors que les entreprises de communication visuelle étaient auparavant génératrices d'emplois. Par contre, l'internet permet à certains professionnels de s'installer à leur compte. C'est le cas pour 30% des salariés ayant perdu leur emploi suite à la restructuration de leur entreprise.

Heinz-Uwe Rübenach, président de l'Association allemande des directeurs de journaux, pense pour sa part que l'emploi n'est pas menacé. Les rédactions locales sont depuis longtemps reliées aux rédactions centrales et les journalistes produisent eux-mêmes des pages prêtes à l'impression. D'après lui, les services en ligne créent de nouveaux emplois, et «les journalistes et autres employés des journaux devront adapter leur activité aux nouvelles technologies afin de faire face à la concurrence et de préserver leur emploi.» Selon une enquête menée auprès des membres de l'Association européenne des directeurs de journaux, les services en ligne recrutent des journalistes ne provenant pas des services de presse classiques, et aucun poste n'a été supprimé dans des entreprises de presse suite à la création de leurs services en ligne.

Selon Michel Muller, secrétaire général de la FILPAC (Fédération des industries du livre, du papier et de la communication), les industries graphiques françaises ont perdu 20.000 emplois en dix ans (1987-1996), avec des effectifs qui sont passés de 110.000 à 90.000 salariés. Les entreprises doivent mettre sur pied des plans sociaux très coûteux pour favoriser le reclassement des personnes licenciées, en créant des emplois souvent artificiels, alors qu'il aurait mieux fallu financer des études fiables sur la manière d'équilibrer créations et suppressions d'emplois lorsqu'il était encore temps.

Peter Leisink, professeur associé en sciences sociales à l'Université d'Utrecht (Pays-Bas), explique que, dans le monde de l'édition, la rédaction des textes et la correction des épreuves sont maintenant des tâches effectuées à domicile, le plus souvent par des travailleurs ayant pris le statut d'indépendants à la suite de licenciements et de délocalisations ou fusions d'entreprises. «Or cette forme d'emploi tient plus du travail précaire que du travail indépendant, car ces personnes n'ont que peu d'autonomie et sont généralement tributaires d'une seule maison d'édition.»

Walter Durling, directeur du géant des télécommunications AT&T aux États-Unis, insiste sur le fait que les nouvelles technologies n'apporteront pas de changement significatif à la situation des salariés au sein de leur entreprise. L'invention du film n'a pas tué le théâtre et celle de la télévision n'a pas fait disparaître le cinéma. Selon lui, les entreprises devraient créer des emplois liés aux nouvelles technologies et les proposer aux salariés dont les postes sont devenus obsolètes. Des arguments bien théoriques alors qu'il s'agit plutôt d'un problème de pourcentage. Combien de créations de postes pour combien de licenciements?

À part quelques cas particuliers mis en avant par les organisations d'employeurs, on observe surtout des suppressions massives d'emplois. Partout dans le monde, des postes à faible qualification technique sont remplacés par des postes demandant des qualifications techniques

élevées. Les salariés peu qualifiés sont licenciés. Certains suivent une formation professionnelle complémentaire, parfois auto-financée sur leur temps libre, et cette formation ne garantit pas pour autant le réemploi.

De leur côté, les syndicats préconisent la création d'emplois par l'investissement, l'innovation, la formation aux nouvelles technologies, la reconversion des travailleurs dont les emplois sont supprimés, des conventions collectives équitables, la défense du droit d'auteur sur l'internet, une meilleure protection des travailleurs dans le secteur artistique, et enfin la défense des télétravailleurs en tant que travailleurs à part entière.

Malgré tous les efforts des syndicats, la situation deviendra-elle aussi dramatique que celle décrite dans une note des actes du colloque? «Certains craignent un futur dans lequel les individus seront forcés de lutter pour survivre dans une jungle électronique. Les mécanismes de survie établis au cours des dernières décennies — tels que relations de travail relativement stables, conventions collectives, représentation des salariés, formation professionnelle procurée par les employeurs et régimes de sécurité sociale co-financés par employeurs et employés — risquent d'être mis à rude épreuve dans un monde du travail qui franchit les frontières à la vitesse de la lumière.»

Le droit d'auteur

Lors du même colloque, Bernie Lunzer, secrétaire-trésorier de la Newspaper Guild (Guilde des journaux) aux États-Unis, insiste sur les batailles juridiques faisant rage sur le non-respect de la propriété intellectuelle sur l'internet. Certains directeurs de publication demandent aux journalistes indépendants (travaillant pour eux) de signer des contrats particulièrement choquants cédant tous leurs droits à la publication en question, avec une contrepartie financière ridicule.

Heinz-Uwe Rübenach, directeur de l'Association allemande de directeurs de journaux, insiste quant à lui sur la nécessité pour les entreprises de presse de contrôler l'utilisation sur le web des articles de leurs journalistes, en demandant une contrepartie financière significative qui leur permettrait d'avoir des fonds pour continuer à investir dans les nouvelles technologies.

Chose inquiétante à l'heure d'une société dite de l'information, le domaine public se réduit comme peau de chagrin. Le Digital Millennium Copyright Act (DMCA) du 27 octobre 1998 porte un coup très rude aux collections numériques américaines, dont le Projet Gutenberg, en plein essor avec le développement du web. De nombreuses oeuvres censées tomber dans le domaine public restent finalement sous copyright.

John Mark Ockerbloom, créateur de The Online Books Page (La page des livres en ligne) en 1993, explique en août 1999 dans un entretien par courriel: «À mon avis, il est important que les usagers comprennent que le copyright est un contrat social conçu pour le bien public — incluant à la fois les auteurs et les lecteurs. Ceci signifie que les auteurs doivent avoir le droit d'utiliser de manière exclusive et pour un temps limité les oeuvres qu'ils ont créées, comme ceci

est spécifié dans la loi actuelle sur le copyright. Mais ceci signifie aussi que les lecteurs ont le droit de copier et de réutiliser ce travail autant qu'ils le veulent à l'expiration de ce copyright. Aux États-Unis, on voit maintenant diverses tentatives visant à retirer ces droits aux lecteurs, en limitant les règles relatives à l'utilisation de ces oeuvres, en prolongeant la durée du copyright (y compris avec certaines propositions visant à le rendre permanent) et en étendant la propriété intellectuelle à des travaux distincts des oeuvres de création (comme on en trouve dans les propositions de copyright pour les bases de données).»

Ce durcissement du copyright affecte tous les pays membres de l'OMPI (Organisation mondiale de la propriété intellectuelle) ayant signé en décembre 1996 le Traité de l'OMPI sur le droit d'auteur, qui vise surtout les éditions numériques, de plus en plus nombreuses.

Dans l'Union européenne, la nouvelle Directive européenne sur le droit d'auteur — appelée aussi directive EUCD (European Union Copyright Directive) — est entérinée en mai 2001 pour harmoniser les législations nationales et fait passer le copyright de 50 ans à 70 ans après le décès de l'auteur. Le nom complet de cette directive est «Directive 2001/29/EC du Parlement européen et du Conseil européen sur l'harmonisation de certains aspects du droit d'auteur et des droits voisins dans la société de l'information». Comme aux États-Unis, elle répond sans nul doute aux pressions exercées par les multinationales culturelles en vue d'une mondialisation du marché, le but officiel étant de renforcer le respect du droit d'auteur sur l'internet et de contrer ainsi le piratage. Cette directive entre progressivement en vigueur dans tous les pays de l'Union européenne avec la mise en place de législations nationales. En France par exemple, la loi DADVSI (Droit d'auteur et droits voisins dans la société de l'information) est promulguée en août 2006, et n'est pas sans susciter de nombreux remous.

Le droit d'auteur adapté au web

Qu'en pensent les auteurs? Jacques Gauchey, journaliste et spécialiste des technologies de l'information, écrit en juillet 1999: «Le droit d'auteur dans son contexte traditionnel n'existe plus. Les auteurs ont besoin de s'adapter à un nouveau paradigme, celui de la liberté totale du flot de l'information. Le contenu original est comme une empreinte digitale: il est incopiable. Il survivra et prospérera donc.»

Alain Bron, consultant en systèmes d'information et écrivain, écrit en novembre 1999: «Je considère aujourd'hui le web comme un domaine public. Cela veut dire que la notion de droit d'auteur sur ce média disparaît de facto: tout le monde peut reproduire tout le monde. La création s'expose donc à la copie immédiate si les copyrights ne sont pas déposés dans les formes usuelles et si les oeuvres sont exposées sans procédures de revenus.»

Selon Xavier Malbreil, auteur multimédia interviewé en mars 2001: «Il y a deux choses. Le web ne doit pas être un espace de non-droit, et c'est un principe qui doit s'appliquer à tout, et notamment au droit d'auteur. Toute utilisation commerciale d'une oeuvre doit ouvrir droit à rétribution. Mais également, le web est un lieu de partage. Échanger entre amis des passages

d’un texte qui vous a plu, comme on peut recopier des passages d’un livre particulièrement apprécié, pour le faire aimer, cela ne peut faire que du bien aux oeuvres, et aux auteurs. La littérature souffre surtout de ne pas être diffusée. Tout ce qui peut concourir à la faire sortir de son ghetto sera positif.»

Nombreux sont les auteurs qui souhaitent prendre en compte la vocation première du web, réseau de création et de diffusion à l’échelon mondial. De ce fait, les adeptes de contrats flexibles comme le copyleft, la licence GPL et les licences Creative Commons sont de plus en plus nombreux.

L’idée du copyleft est lancée dès 1984 par Richard Stallman, développeur de logiciels, qui fonde la Free Software Foundation (FSF — Fondation pour le logiciel libre) dans ce but. Conçu à l’origine pour les logiciels, le copyleft est formalisé par la GPL (General Public License — Licence publique générale) et étendu par la suite à toute oeuvre de création. Il contient la déclaration normale du copyright affirmant le droit d’auteur, mais son originalité est de donner à l’usager le droit de librement redistribuer l’oeuvre et de la modifier. L’usager s’engage toutefois à ne revendiquer ni le travail original ni les oeuvres dérivées, qui sont elles aussi placées d’emblée sous licence GPL.

La GPL se décline en plusieurs variantes. La licence utilisée pour la documentation accompagnant les logiciels libres — manuels, livres et autres documents écrits — est la GFDL (GNU Free Documentation License — Licence de documentation libre GNU). La GFDL est utilisée ensuite pour toute oeuvre documentaire, notamment pour les dictionnaires et les encyclopédies en ligne. C’est par exemple la licence utilisée pour les articles de Wikipédia avant l’utilisation d’une licence Creative Commons.

Le principe de la licence Creative Commons (CC) est développé en 2001 par Lawrence Lessig, juriste, professeur de droit à la Stanford Law School (Université de droit de Stanford) en Californie et ardent défenseur d’un internet créatif sur lequel les oeuvres pourraient être non seulement diffusées mais aussi réutilisées lorsque les auteurs donnent leur accord.

L’organisme du même nom lance en décembre 2002 ses premières licences-type, qui sont des contrats flexibles de droit d’auteur compatibles avec une diffusion sur l’internet, valables pour tout type de création (texte, photo, film, musique, site web, etc.). Ces licences-type donnent l’autorisation (ou non) de copier, distribuer, communiquer, remixer ou transformer l’oeuvre originale. Rédigées par des juristes, elles sont accompagnées d’un résumé court accessible à tous.

Six licences sont disponibles: (1) la licence CC BY (la plus large, qui requiert seulement la mention de l’auteur, du titre et de la source en cas de réutilisation de l’oeuvre), (2) la licence CC BY-SA (qui requiert la même licence en cas de réutilisation de l’oeuvre), (3) la licence CC BY-NC (qui interdit les oeuvres commerciales en cas de réutilisation de l’oeuvre), (4) la licence CC BY-ND (qui interdit les oeuvres dérivées en cas de réutilisation de l’oeuvre), (5) la licence CC BY-NC-SA (qui interdit les oeuvres commerciales et requiert la même licence en cas de réutilisation de l’oeuvre), (6) la licence CC BY-NC-ND (qui interdit les oeuvres commerciales et les oeuvres dérivées en cas de réutilisation de l’oeuvre).

Après une version 1.0 (publiée en décembre 2002) suivie d'une version 2.0 (publiée en mai 2004), la version 3.0 (publiée en février 2007) instaure une licence internationale et la compatibilité avec d'autres licences similaires, dont le copyleft, la licence GPL (General Public License — Licence publique générale) et la licence GFDL (GNU Free Documentation License — Licence de documentation libre GNU). Suite à un sommet mondial organisé en septembre 2011 auquel s'ajoutent deux années de discussions publiques documentées sur le web, Creative Commons publie en novembre 2013 la version 4.0 de ses six licences.

Qui utilise une licence Creative Commons? Le premier éditeur à utiliser une licence Creative Commons est O'Reilly Media. Une licence Creative Commons est utilisée aussi par Wikipédia pour tous ses articles, rédigés par des milliers de contributeurs dans des dizaines de langues différentes, auxquels s'ajoutent tous ceux qui fournissent des illustrations (images, photos, dessins, cartes, graphiques, etc.).

Suivent ensuite des milliers d'auteurs, illustrateurs, photographes, cinéastes, musiciens, éducateurs, chercheurs et autres créateurs partageant leurs travaux sur la toile. Une licence Creative Commons est utilisée pour un million d'oeuvres en 2003, 4,7 millions d'oeuvres en 2004, 20 millions d'oeuvres en 2005, 50 millions d'oeuvres en 2006, 90 millions d'oeuvres en 2007, 130 millions d'oeuvres en 2008, 400 millions d'oeuvres en 2010 et 882 millions d'oeuvres en 2014.

Les librairies

Basée au Royaume-Uni, l'Internet Bookshop (IBS) se trouve être la plus grande librairie en ligne européenne en 1997, avec un catalogue de 1,4 million de titres. Elle lance des initiatives originales et inédites qui sont reprises par Amazon aux États-Unis dans les mois qui suivent.

L'Internet Bookshop développe d'abord en janvier 1997 un système de partenariat sur le web. Tout possesseur d'un site web peut devenir partenaire de l'Internet Bookshop en sélectionnant sur son propre site un certain nombre de titres présents dans le catalogue de la librairie. Celle-ci prend en charge toute la partie commerciale, à savoir les commandes, les envois et les factures. L'usager partenaire reçoit 10% du prix des ventes. C'est la première fois qu'une librairie en ligne propose une part des bénéfices par le biais du web, initiative reprise par Amazon au printemps de la même année avec sa politique de sites affiliés (dénommés «associés»), entraînant à terme la nécessité d'une nouvelle réglementation dans ce domaine.

Une autre initiative originale qui débute en octobre 1997 est une politique de grosses remises, chose inconnue jusque-là. L'Internet Bookshop propose des remises allant jusqu'à 45% pour certains best-sellers, prenant le risque d'une guerre des prix et des droits avec les éditeurs et les libraires traditionnels. L'idée est ensuite reprise outre-Atlantique par Amazon et Barnes & Noble.com, qui se lancent dans une guerre des prix entre eux.

En octobre 1997, l’Internet Bookshop attend également la réaction des éditeurs et des libraires traditionnels à sa décision de vendre des livres provenant des États-Unis, une initiative débutée un mois auparavant. Une deuxième librairie en ligne britannique, Waterstones, annonce elle aussi son intention d’introduire des titres américains dans son catalogue à compter de janvier 1998. The Publishers Association (Association des éditeurs), organisme représentant les éditeurs au Royaume-Uni, a fort à faire pour étudier les doléances des ces derniers, jointes à celles des libraires traditionnels, qui souhaiteraient non seulement faire interdire la vente de titres américains par des librairies en ligne britanniques, mais aussi faire interdire au Royaume-Uni l’activité des librairies en ligne américaines, à savoir qu’elles ne puissent pas vendre de livres à des clients britanniques.

En 1997 et 1998, la rubrique IBS News de l’Internet Bookshop permet de suivre pas à pas le combat engagé par les librairies en ligne contre les associations d’éditeurs et de libraires traditionnels, le but étant d’obtenir la suppression totale des frontières pour la vente des livres. Comme on le voit, ce qui nous paraît évident maintenant ne l’était guère à l’époque. Mais, de par la structure même de l’internet, l’abolition des frontières dans le marché du livre est inévitable, tout comme l’instauration d’une législation adéquate. D’autres librairies en ligne ne tardent pas à suivre l’exemple de l’Internet Bookshop, à savoir vendre des livres par-delà les frontières.

Concernant la fiscalité — autre point sensible — un accord-cadre entre les États-Unis et l’Union européenne est conclu en décembre 1997. L’internet est désormais considéré comme une zone de libre-échange pour les logiciels, les films et les livres achetés sur l’internet. Les biens matériels sont soumis au régime existant dans les pays concernés, avec perception de la TVA sans frais de douane supplémentaires. Cet accord-cadre est suivi d’une convention internationale.

En 2000, les grandes librairies en ligne françaises sont Fnac.com, Alapage, Chapitre.com et BOL.fr. Après avoir créé ses deux premières filiales européennes au Royaume-Uni et en Allemagne en octobre 1998, Amazon lance sa troisième filiale européenne en France en août 2000 avec livres, musique, DVD et vidéos (puis logiciels et jeux vidéo en juin 2001), tout comme une livraison en 48 heures. À cette date, la vente de livres en ligne en France ne représente que 0,5% du marché du livre, alors que ce pourcentage est de 5,4% aux États-Unis.

Les librairies en ligne françaises ne peuvent pas se permettre les réductions substantielles proposées par les librairies en ligne basées au Royaume-Uni et en Amérique du Nord, pays dans lesquels le prix du livre est libre. Si la loi française sur le prix unique du livre — dénommée loi Lang, du nom du ministre de la culture à l’origine de cette loi — laisse peu de latitude, à savoir un rabais de 5% maximum sur ce prix, les librairies en ligne restent optimistes sur les perspectives d’un marché en ligne, un optimisme qui sera confirmé plus tard.

Qu’en est-il des librairies indépendantes, frappées de plein fouet par les librairies en ligne? Certaines résistent, comme la librairie Ulysse, fondée en 1971 par Catherine Domain et consacrée aux voyages. Nichée au coeur de Paris, dans l’île Saint-Louis, la librairie propose 20.000 livres, cartes et revues neufs et d’occasion, qui recèlent des trésors introuvables ailleurs. Catherine Domain est membre du Syndicat national de la librairie ancienne et moderne

(SLAM), du Club des explorateurs et du Club international des grands voyageurs. À la fois libraire et grande voyageuse, elle voyage tous les étés sur l'une des mers du globe pendant que son compagnon, expert en cartes anciennes, tient la librairie.

En janvier 1999, Catherine Domain décide de se lancer dans un voyage autrement plus ingrat, virtuel cette fois, à savoir la réalisation d'un site web en autodidacte alors que ses connaissances en informatique sont très sommaires, comme beaucoup d'entre nous à l'époque. Elle raconte en décembre 1999 lors d'un entretien par courriel: «Mon site est embryonnaire et en construction. Il se veut à l'image de ma librairie, un lieu de rencontre avant d'être un lieu commercial. Il sera toujours en perpétuel devenir! L'internet me prend la tête, me bouffe mon temps et ne me rapporte presque rien, mais cela ne m'ennuie pas…» Elle est toutefois pessimiste sur l'avenir des librairies comme la sienne. «Internet tue les librairies spécialisées. En attendant d'être dévorée, je l'utilise comme un moyen d'attirer les clients chez moi, et aussi de trouver des livres pour ceux qui n'ont pas encore l'internet chez eux! Mais j'ai peu d'espoir…»

En 2005, elle crée une deuxième librairie de voyage à Hendaye, sur la côte sud de l'Atlantique, non loin de la frontière espagnole, avec une vue imprenable sur la plage et l'océan. Ouverte du 20 juin au 20 septembre, avec des horaires qui fluctuent en fonction des marées et de la météo, la librairie est située dans un bâtiment mauresque classé monument historique, qui se trouve être l'ancien casino. À marée haute, la librairie est «comme un paquebot de livres qui va prendre la mer, qu'elle prend quelquefois d'ailleurs».

Dix ans après la création de son site web, Catherine Domain est beaucoup moins critique à l'égard de l'internet, qui lui permet de lancer sa propre maison d'édition pour publier des livres de voyage. Elle écrit en avril 2010: «L'internet a pris de plus en plus de place dans ma vie! Il me permet d'être éditeur grâce à de laborieuses formations Photoshop, InDesign et autres. C'est une grande joie de constater que la volonté politique de garder le pékin devant son ordinateur afin qu'il ne fasse pas la révolution peut être mise en échec par des apéros géants et spontanés de milliers de personnes [organisés via Facebook] qui veulent se voir et se parler en vrai. Décidément il y aura toujours des rebondissements inattendus aux inventions, entre autres. Quand j'ai commencé à utiliser l'internet, je ne m'attendais vraiment pas à devenir éditeur.»

Les librairies numériques

Des librairies d'un genre nouveau se spécialisent d'emblée dans la vente de livres numériques, par exemple Numilog, créé en décembre 1999, ou Mobipocket, créé en mars 2000. Au fil des ans, après avoir lu ces livres sur l'écran de son ordinateur (portable ou non), on peut aussi les lire sur des supports mobiles (PDA, smartphone, liseuse, tablette).

Le premier format de livre numérique est le PDF (Portable Document Format), lancé en juin 1993 par la société Adobe en même temps que l'Acrobat Reader (gratuit, pour lire les PDF) et l' Adobe Acrobat (payant, pour créer les PDF). Ce format fige un document afin de conserver

son contenu et sa présentation quelle que soit la plateforme utilisée pour créer le PDF puis pour le lire. Beaucoup plus tard, après avoir été un format propriétaire, le format PDF devient un standard ouvert en juillet 2008, tout comme une norme ISO (ISO 32000-1:2008).

La fin des années 1990 est marquée par la prolifération des formats, chacun lançant son propre format de livre numérique dans le cadre d'un marché naissant promis à une expansion rapide. L'industrie du livre numérique promeut en septembre 1999 un format commun, l'OeB (Open eBook), basé sur le langage XML et destiné à normaliser le contenu, la structure et la présentation des livres numériques. Le format OeB sert de base à de nombreux formats, par exemple le format LIT (Microsoft) ou le format PRC (Mobipocket), avant d'être remplacé par le format EPUB en septembre 2007. D'autres formats numériques sont développés en parallèle pour les livres braille et audio.

Selon Denis Zwirn, fondateur et président de la librairie numérique Numilog, interviewé en août 2007: «Le livre numérique est un produit commercial et un outil au service de la lecture. Il suffit de proposer des textes lisibles facilement sur les supports de lecture électronique variés qu'utilisent les gens, l'encre électronique pouvant progressivement envahir tous ces supports. Et de les proposer de manière industrielle. Le livre numérique est en train de devenir un produit de masse, riche de formes multiples comme l'est le livre traditionnel.» Le catalogue de Numilog comprend 3.500 titres (livres et revues) en 2003, 35.000 titres en décembre 2006 et 50.000 titres en janvier 2009, suite à des partenariats avec une centaine d'éditeurs francophones et anglophones, auxquels s'ajoutent des services spécifiques pour les bibliothèques et les librairies.

Une autre grande librairie numérique est celle de Mobipocket. Fondée en mars 2000 par Thierry Brethes et Nathalie Ting, la société Mobipocket se spécialise d'emblée dans la lecture et la distribution sécurisée de livres pour PDA. Son logiciel de lecture, le Mobipocket Reader, est utilisable sur n'importe quel PDA, puis sur ordinateur en avril 2002 et sur smartphone ensuite. La librairie de Mobipocket comprend 6.000 titres dans plusieurs langues (français, anglais, allemand, espagnol) en 2003, et ces livres sont distribués dans la librairie de Mobipocket et dans des librairies partenaires. Mobipocket est racheté par Amazon en avril 2005 dans l'optique du lancement du Kindle deux ans plus tard, et les 39.000 titres de la librairie de Mobipocket sont intégrés au Kindle Store en avril 2007.

L'EPUB supplante peu à peu le PDF. Nicolas Pewny, consultant en édition électronique, précise en octobre 2015 dans un entretien par courriel: «Les PDF peuvent contenir des liens et des boutons, des champs de formulaire, des contenus audio et vidéo. Ils prennent également en charge les signatures électroniques. Le principal avantage de ce format est d'être parfaitement fidèle à la mise en page d'origine. Par contre, il est assez lourd et manque de souplesse; il ne s'adapte que très peu à l'écran utilisé. Totalement différent du PDF, le format EPUB permet d'afficher un texte ajusté au tout dernier moment en fonction de l'écran et des choix du lecteur au moment précis de la lecture. La dernière version standardisée, EPUB3, repose sur HTML5, la dernière révision majeure d'HTML, format de données conçu pour représenter les pages web. Le format EPUB3 peut interroger les fonctionnalités du lecteur via JavaScript. Il peut ainsi adapter le contenu avec une grande souplesse. À mon avis, EPUB3 est le meilleur format pour la publication électronique aujourd'hui. Plusieurs applications permettent de créer un fichier EPUB directement, dont des logiciels libres. La plupart des liseuses sont compatibles avec le

format EPUB, excepté le Kindle d’Amazon, qui nécessite une transformation (par exemple avec le logiciel libre Calibre). Diverses applications de lecture supportant l’EPUB sont également disponibles pour les ordinateurs personnels, les tablettes tactiles et les smartphones.»

Marc Autret, développeur et infographiste, explique pour sa part en juin 2011: «Je déplore que l’émergence de l’EPUB ait provoqué l’anéantissement pur et simple du PDF comme format de livre numérique. Le fait que les éléments d’interactivité disponibles au sein du PDF ne soient pas supportés par les plateformes nomades actuelles a aboli toute possibilité d’expérimentation dans cette voie, qui m’apparaissait extrêmement prometteuse. Alors que l’édition imprimée fait la place à des objets de nature très différente, entre le livre d’art de très haute facture et le livre “tout terrain”, le marché de l’e-book s’est développé d’emblée sur un mode totalitaire et ségrégationniste, comparable en cela à une guerre de systèmes d’exploitation plutôt qu’à une émulation technique et culturelle. De fait, il existe fort peu de livres numériques PDF tirant parti des possibilités de ce format.»

«Dans l’inconscient collectif, le PDF reste une sorte de duplicata statique de l’ouvrage imprimé et personne ne veut lui voir d’autre destin. L’EPUB, qui n’est rien d’autre qu’une combinaison XHTML/CSS (certes avec des perspectives JavaScript), consiste à mettre le livre numérique “au pas” du web. C’est une technologie très favorable aux contenus structurés, mais très défavorable à l’artisanat typographique. Elle introduit une vision étroite de l’oeuvre numérique, réduite à un flux d’information. On ne le mesure pas encore, mais la pire catastrophe culturelle de ces dernières décennies est l’avènement du XML, ce langage qui précalibre et contamine notre façon de penser les hiérarchies. Le XML et ses avatars achèvent de nous enfermer dans les invariants culturels occidentaux.»

Les auteurs

Le deuxième roman d’Alain Bron, «Sanguine sur toile», est publié sous forme imprimée en 1999 par les éditions du Choucas et sous forme numérique (PDF) en 2000 par les éditions 00h00. Dans ce roman, l’internet est un personnage en soi, comme expliqué par Alain Bron en novembre 1999 dans un entretien par courriel: «Plutôt que de le décrire dans sa complexité technique, le réseau est montré comme un être tantôt menaçant, tantôt prévenant, maniant parfois l’humour. N’oublions pas que l’écran d’ordinateur joue son double rôle: il montre et il cache. C’est cette ambivalence qui fait l’intrigue du début à la fin. Dans ce jeu, le grand gagnant est bien sûr celui ou celle qui sait s’affranchir de l’emprise de l’outil pour mettre l’humanisme et l’intelligence au-dessus de tout.»

Quelle est exactement l’histoire? «La “toile”, c’est celle du peintre, c’est aussi l’autre nom d’internet: le web — la toile d’araignée. “Sanguine” évoque le dessin et la mort brutale. Mais l’amour des couleurs justifierait-il le meurtre? “Sanguine sur toile” évoque l’histoire singulière d’un internaute pris dans la tourmente de son propre ordinateur, manipulé à distance par un très mystérieux correspondant qui n’a que vengeance en tête. J’ai voulu emporter le lecteur dans les univers de la peinture et de l’entreprise, univers qui s’entrelacent, s’échappent, puis se

rejoignent dans la fulgurance des logiciels. Le lecteur est ainsi invité à prendre l'enquête à son propre compte pour tenter de démêler les fils tressés par la seule passion. Pour percer le mystère, il devra répondre à de multiples questions. Le monde au bout des doigts, l'internaute n'est-il pas pour autant l'être le plus seul au monde? Compétitivité oblige, jusqu'où l'entreprise d'aujourd'hui peut-elle aller dans la violence? La peinture tend-t-elle à reproduire le monde ou bien à en créer un autre? Enfin, j'ai voulu montrer que les images ne sont pas si sages. On peut s'en servir pour agir, voire pour tuer.»

À la fois consultant en systèmes d'information et écrivain, Alain Bron relate aussi son parcours: «J'ai passé une vingtaine d'années chez Bull. Là, j'ai participé à toutes les aventures de l'ordinateur et des télécommunications, j'ai été représentant des industries informatiques à l'ISO [Organisation internationale de normalisation], et chairman du groupe réseaux du consortium X/Open. J'ai connu aussi les tout débuts d'internet avec mes collègues de Honeywell aux États-Unis (fin 1978). Je suis actuellement consultant en systèmes d'information où je m'occupe de la bonne marche de grands projets informatiques. Et j'écris. J'écris depuis mon adolescence. Des nouvelles (plus d'une centaine), des essais psychosociologiques ("La gourmandise du tapir" et "La démocratie de la solitude"), des articles et des romans. C'est à la fois un besoin et un plaisir jubilatoire.»

La première expérience numérique du romancier espagnol Arturo Pérez-Reverte est un peu différente. Sa série best-seller relate les aventures du capitaine Alatriste au 17e siècle, avec trois titres parus en 1996, 1997 et 1998. Le nouveau titre à paraître fin 2000 s'intitule «El Oro del Rey» («L'Or du Roi»). En novembre 2000, en collaboration avec son éditeur Alfaguara, l'auteur décide de vendre ce nouveau titre sous forme numérique sur un page spécifique du portail web Inicia, en exclusivité pendant un mois, avant sa sortie en librairie. Le roman est disponible au format PDF pour 2,90 euros, un prix très inférieur aux 15,10 euros annoncés pour le livre imprimé.

Résultat de l'expérience, le nombre de téléchargements est très satisfaisant, mais pas celui des paiements. Un mois après la mise en ligne du PDF, on compte 332.000 téléchargements, avec paiement par 12.000 lecteurs seulement. Marilo Ruiz de Elvira, directrice de contenus du portail Inicia, explique dans un communiqué: «Pour tout acheteur du livre numérique, il y avait une clé pour le télécharger en 48 heures sur le site internet et, surtout au début, de nombreux usagers ont échangé ce code d'accès dans les forums de chats et ont téléchargé leur exemplaire sans payer. On a voulu faire un test et cela faisait partie du jeu. Arturo Pérez-Reverte voulait surtout qu'on le lise.»

Un cinquième tome imprimé paraît fin 2003. Devenue un succès planétaire avec 4 millions d'exemplaires vendus depuis ses débuts, la saga donne également naissance en 2006 au film Alatriste, une superproduction espagnole de 20 millions d'euros. Un sixième tome imprimé sort fin 2006, suivi d'un septième tome fin 2011.

Une autre expérience numérique est celle du romancier brésilien Paulo Coelho, devenu mondialement célèbre après la parution de «L'Alchimiste». Début 2003, ses livres, traduits en 56 langues, ont été vendus en 53 millions d'exemplaires dans 155 pays, dont 6,5 millions d'exemplaires dans les pays francophones. En mars 2003, l'auteur décide de distribuer plusieurs de ses romans gratuitement en version PDF, en diverses langues, avec l'accord de ses éditeurs

respectifs, dont Anne Carrière, son éditrice en France. Trois romans sont disponibles en français: «Manuel du guerrier de la lumière», «La cinquième montagne» et «Veronika décide de mourir».

Pourquoi une telle décision? L'auteur déclare à la même date par le biais de son éditrice: «Comme le français est présent, à plus ou moins grande échelle, dans le monde entier, je recevais sans cesse des courriers électroniques d'universités et de personnes habitant loin de la France, qui ne trouvaient pas mes oeuvres.»

À la question classique relative au préjudice éventuel sur les ventes futures, l'auteur répond: «Seule une minorité de gens a accès à l'internet, et le livre au format e-book ne remplacera jamais le livre papier.» Une remarque très juste en 2003, mais qui n'est peut-être plus de mise ensuite. Paulo Coelho réitère toutefois l'expérience au printemps 2011 avec de nouveaux titres, pour la plus grande joie de ses lecteurs.

Jean-Paul, écrivain hypermédia et webmestre du site Cotres.net, propose depuis octobre 1998 de beaux parcours littéraires utilisant l'hyperlien. Il relate en juin 2000: «La navigation par hyperliens se fait en rayon (j'ai un centre d'intérêt et je clique méthodiquement sur tous les liens qui s'y rapportent) ou en louvoiements (de clic en clic, à mesure qu'ils apparaissent, au risque de perdre de vue mon sujet). Bien sûr, les deux sont possibles avec l'imprimé. Mais la différence saute aux yeux: feuilleter n'est pas cliquer. L'internet a donc changé mon rapport à l'écriture. On n'écrit pas de la même manière pour un site que pour un scénario, une pièce de théâtre, etc. Depuis, j'écris (compose, mets en page, en scène) directement à l'écran. L'état "imprimé" de mon travail n'est pas le stade final, le but; mais une forme parmi d'autres, qui privilégie la linéarité et l'image, et qui exclut le son et les images animées.»

«C'est finalement dans la publication en ligne (l'entoilage?) que j'ai trouvé la mobilité, la fluidité que je cherchais. Le maître mot y est "chantier en cours", sans palissades. Accouchement permanent, à vue, comme le monde sous nos yeux. Provisoire, comme la vie qui tâtonne, se cherche, se déprend, se reprend. Avec évidemment le risque souligné par les gutenbergs, les orphelins de la civilisation du livre: plus rien n'est sûr. Il n'y a plus de source fiable, elles sont trop nombreuses, et il devient difficile de distinguer un clerc d'un gourou. Mais c'est un problème qui concerne le contrôle de l'information. Pas la transmission des émotions.»

«En fait, ce n'est pas sur la toile, c'est dans le premier Mac que j'ai découvert l'hypermédia à travers l'auto-apprentissage d'HyperCard. Je me souviens encore de la stupeur dans laquelle j'ai été plongé, durant le mois qu'a duré mon apprentissage des notions de boutons, liens, navigation par analogies, par images, par objets. L'idée qu'un simple clic sur une zone de l'écran permettait d'ouvrir un éventail de piles de cartes dont chacune pouvait offrir de nouveaux boutons dont chacun ouvrait un nouvel éventail dont… bref l'apprentissage de tout ce qui aujourd'hui sur la toile est d'une banalité de base, cela m'a fait l'effet d'un coup de foudre (il paraît que Steve Jobs et son équipe eurent le même choc lorsqu'ils découvrirent l'ancêtre du Mac dans les laboratoires de Rank Xerox). Depuis, j'écris (compose, mets en page, en scène) directement à l'écran. L'état "imprimé" de mon travail n'est pas le stade final, le but; mais une forme parmi d'autres, qui privilégie la linéarité et l'image, et qui exclut le son et les images animées.»

Les bibliothèques

La première bibliothèque publique à avoir un site web est la Bibliothèque municipale d'Helsinki (Finlande), qui lance son site en février 1994.

Selon le document «Internet and the Library Sphere» (Internet et la sphère des bibliothèques) publié en ligne sur le site de la Commission européenne, environ 1.000 bibliothèques publiques dans 26 pays disposent d'un site web en novembre 1998. Les sites sont hétérogènes. Certaines bibliothèques se contentent de mentionner leur adresse postale et leurs heures d'ouverture, tandis que d'autres proposent toute une gamme de services, y compris un accès direct à leur catalogue en ligne. Les pays les plus représentés sont la Finlande (247 bibliothèques), la Suède (132 bibliothèques), le Royaume-Uni (112 bibliothèques), le Danemark (107 bibliothèques), l'Allemagne (102 bibliothèques), les Pays-Bas (72 bibliothèques), la Lituanie (51 bibliothèques), l'Espagne (56 bibliothèques) et la Norvège (45 bibliothèques). La Russie propose un site commun pour 26 bibliothèques publiques de recherche. Les pays nouvellement représentés fin 1998 sont la République tchèque (29 bibliothèques) et le Portugal (3 bibliothèques).

Lancé par la CENL (Conference of European National Librarians — Conférence des bibliothécaires nationaux européens) en janvier 1997, Gabriel est un portail trilingue offrant un point d'accès commun aux services internet des bibliothèques nationales européennes.

Gabriel est l'acronyme de «Gateway and Bridge to Europe's National Libraries» (Passerelle et pont vers les bibliothèques nationales européennes). On lit aussi sur le site que, outre son universalité, ce nom «rappelle également les travaux de Gabriel Naudé, dont L'"Advis pour dresser une bibliothèque" (Paris, 1627) est le premier travail théorique en Europe sur les bibliothèques et constitue ainsi un point de départ sur les bibliothèques de recherche modernes. Le nom Gabriel est aussi employé dans de nombreuses langues européennes et vient de l'Ancien Testament, Gabriel étant l'un des archanges, ou messager céleste. Il est également présent dans le Nouveau Testament et dans le Coran.»

Le site offre en 1998 des liens vers les services internet des 38 bibliothèques nationales participantes (Allemagne, Autriche, Belgique, Bulgarie, Danemark, Espagne, Estonie, Finlande, France, Grèce, Hongrie, Irlande, Islande, Italie, Lettonie, Liechtenstein, Lituanie, Luxembourg, Macédoine, Malte, Norvège, Pays-Bas, Pologne, Portugal, République slovaque. République tchèque, Roumanie, Royaume-Uni, San Marino, Suède, Suisse, Turquie, Vatican). Les services internet sont très divers d'une bibliothèque à l'autre, avec une liste complète par bibliothèque. Ces services sont par exemple des catalogues en ligne, des bibliographies nationales, des catalogues collectifs nationaux, des index de périodiques, des serveurs web et des gophers (menus textuels à plusieurs niveaux). Une rubrique spécifique présente les projets communs à plusieurs pays. La recherche sur Gabriel est possible par pays et par type de service.

Beaucoup plus tard, au cours de l'été 2005, Gabriel fusionne avec le site web de la Bibliothèque européenne pour proposer un portail commun aux 43 bibliothèques nationales. Lancée en janvier 2004, la Bibliothèque européenne est issue du projet TEL (Telematics & Electronic Libraries — Télématique et bibliothèques électroniques), un projet financé par la Commission européenne pendant trois ans (2001-2003) pour étudier la faisabilité d'un service pan-européen donnant accès aux ressources disséminées dans toutes les bibliothèques nationales d'Europe.

Dans un premier temps, la Bibliothèque européenne offre un point d'accès commun aux catalogues de 18 bibliothèques nationales (Allemagne, Autriche, Croatie, Danemark, Estonie, Finlande, France, Italie, Hongrie, Lettonie, Pays-Bas, Portugal, République tchèque, Royaume-Uni, Serbie, Slovaquie, Slovénie, Suisse), avec accès gratuit ou payant aux documents numériques de ces bibliothèques. S'y ajoutent les catalogues de la CENL et de l'ICCU (Instituto Centrale per il Catalogo Unico delle Bibliothece Italiane — Institut central pour le catalogue commun des bibliothèques italiennes). Les catalogues des 25 autres bibliothèques nationales européennes sont intégrés dans une phase ultérieure. Trois ans après la fusion de Gabriel avec la Bibliothèque européenne, Europeana est lancée en novembre 2008 en tant que grande bibliothèque numérique publique européenne.

Les bibliothécaires

Nombre de bibliothécaires deviennent des cyberthécaires et des webmestres, par exemple Peter Raggett à l'OCDE (Organisation de coopération et de développement économiques), Bruno Didier à l'Institut Pasteur de Paris et Bakayoko Bourahima à l'ENSEA (École nationale supérieure de statistique et d'économie appliquée) d'Abidjan.

Peter Raggett est directeur de la bibliothèque de l'OCDE, une organisation internationale regroupant trente pays membres. Au noyau d'origine, constitué des pays d'Europe de l'Ouest et d'Amérique du Nord, viennent s'ajouter le Japon, l'Australie, la Nouvelle-Zélande, la Finlande, le Mexique, la République tchèque, la Hongrie, la Pologne et la Corée. Réservée aux fonctionnaires de l'organisation, la bibliothèque comprend 60.000 monographies et 2.500 périodiques imprimés en 1998, et elle propose aussi une collection de microfilms et de CD-ROM tout comme la consultation de bases de données telles que Dialog, Lexis-Nexis et UnCover. La bibliothèque lance en 1996 ses pages intranet, qui deviennent rapidement une source d'information majeure pour le personnel de l'organisation.

Peter Raggett explique en août 1999 dans un entretien par courriel: «Je dois filtrer l'information pour les usagers de la bibliothèque, ce qui signifie que je dois bien connaître les sites et les liens qu'ils proposent. J'ai sélectionné plusieurs centaines de sites pour en favoriser l'accès à partir de l'intranet de l'OCDE. Cette sélection fait partie du bureau de référence virtuel proposé par la bibliothèque à l'ensemble du personnel. Outre de nombreux liens, ce bureau de référence

contient des pages recensant les articles, monographies et sites web correspondant aux différents projets de recherche en cours à l'OCDE, l'accès en réseau aux CD-ROM et une liste mensuelle des nouveaux livres achetés par la bibliothèque.»

«L'internet offre aux chercheurs un stock d'informations considérable. Le problème pour eux est de trouver ce qu'ils cherchent. Jamais auparavant on n'avait senti une telle surcharge d'informations, comme on la sent maintenant quand on tente de trouver un renseignement sur un sujet précis en utilisant les moteurs de recherche disponibles sur l'internet. Lorsqu'on utilise un moteur de recherche comme Lycos ou AltaVista ou un répertoire comme Yahoo!, on voit vite la difficulté de trouver des sites utiles sur un sujet donné. La recherche fonctionne assez bien sur un sujet très précis, par exemple si on cherche des informations sur une personne au nom inhabituel, mais elle donne un trop grand nombre de résultats si on cherche des informations sur un sujet assez vaste. Par exemple, si on lance une recherche sur le web pour "Russie *et* transport", dans le but de trouver des statistiques sur l'utilisation des trains, des avions et des bus en Russie, les premiers résultats qu'on trouve sont les compagnies de transport de fret qui ont des relations d'affaires avec la Russie.»

«À mon avis, les bibliothécaires auront un rôle important à jouer pour améliorer la recherche et l'organisation de l'information sur le réseau. Je prévois aussi une forte expansion de l'internet pour l'enseignement et la recherche. Les bibliothèques seront amenées à créer des bibliothèques numériques permettant à un étudiant de suivre un cours proposé par une institution à l'autre bout du monde. La tâche du bibliothécaire sera de filtrer les informations pour le public. Personnellement, je me vois de plus en plus devenir un bibliothécaire virtuel. Je n'aurai pas l'occasion de rencontrer les usagers, ils me contacteront plutôt par courriel, par téléphone ou par fax, j'effectuerai la recherche et je leur enverrai les résultats par voie électronique.»

Bruno Didier est bibliothécaire à l'Institut Pasteur de Paris, une fondation privée spécialisée dans la prévention et le traitement des maladies infectieuses (malaria, tuberculose, sida, fièvre jaune, dengue, poliomyélite, etc.), avec plusieurs instituts dans le monde. Séduit par les perspectives qu'offre l'internet pour la recherche documentaire, il crée le site web de la bibliothèque en 1996 et devient son webmestre.

Il explique en août 1999: «Le site web de la bibliothèque a pour vocation principale de servir la communauté pasteurienne. Il est le support d'applications devenues indispensables à la fonction documentaire dans un organisme de cette taille: bases de données bibliographiques, catalogue, commande de documents et bien entendu accès à des périodiques en ligne (un peu plus d'une centaine actuellement). C'est également une vitrine pour nos différents services, en interne mais aussi dans toute la France et à l'étranger. Il tient notamment une place importante dans la coopération documentaire avec les instituts du réseau Pasteur à travers le monde. Enfin j'essaie d'en faire une passerelle adaptée à nos besoins pour la découverte et l'utilisation d'internet. Je développe et maintiens les pages du serveur, ce qui s'accompagne d'une activité de veille régulière. Par ailleurs je suis responsable de la formation des usagers, ce qui se ressent dans mes pages. Le web est un excellent support pour la formation, et la plupart des réflexions actuelles sur la formation des usagers intègrent cet outil.»

L’activité professionnelle de Bruno Didier a changé de manière radicale, tout comme celle de ses collègues. «C’est à la fois dans nos rapports avec l’information et avec les usagers que les changements ont eu lieu. Nous devenons de plus en plus des médiateurs, et peut-être un peu moins des conservateurs. Mon activité actuelle est typique de cette nouvelle situation: d’une part dégager des chemins d’accès rapides à l’information et mettre en place des moyens de communication efficaces, d’autre part former les utilisateurs à ces nouveaux outils. Je crois que l’avenir de notre métier passe par la coopération et l’exploitation des ressources communes. C’est un vieux projet certainement, mais finalement c’est la première fois qu’on dispose enfin des moyens de le mettre en place.»

Bakayoko Bourahima est responsable de la bibliothèque de l’ENSEA (École nationale supérieure de statistique et d’économie appliquée) à Abidjan (Côte d’ivoire). L’ENSEA assure la formation de statisticiens pour les pays africains d’expression française. Le site web de l’ENSEA est mis en ligne en avril 1999 dans le cadre du réseau REFER, un réseau créé par l’Agence universitaire de la francophonie (AUF) pour desservir la communauté scientifique et technique en Afrique, en Asie et en Europe orientale, avec 24 pays participants en 2002.

Bakayoko Bourahima assure à la fois la gestion de l’information et la diffusion des travaux publiés par l’ENSEA. Il explique en juillet 2000: «Le service de la bibliothèque travaille à deux projets d’intégration du web pour améliorer ses prestations. J’espère bientôt pouvoir mettre à la disposition de mes usagers un accès internet pour l’interrogation de bases de données. Par ailleurs, j’ai en projet de réaliser et de mettre sur l’intranet et sur le web un certain nombre de services documentaires (base de données thématique, informations bibliographiques, service de références bibliographiques, bulletin analytique des meilleurs travaux d’étudiants…). Il s’agit donc pour la bibliothèque, si j’obtiens les financements nécessaires pour ces projets, d’utiliser pleinement l’internet pour donner à notre École un plus grand rayonnement et de renforcer sa plateforme de communication avec tous les partenaires possibles. En intégrant cet outil au plan de développement de la bibliothèque, j’espère améliorer la qualité et élargir la gamme de l’information scientifique et technique mise à la disposition des étudiants, des enseignants et des chercheurs, tout en étendant considérablement l’offre des services de la bibliothèque.»

Les bibliothèques numériques

Qu’est-ce exactement qu’une bibliothèque numérique? Selon la British Library, qui tente en 1997 d’en donner une définition sur son site web avant de lancer la sienne, une collection numérique devient une bibliothèque numérique si elle répond aux quatre critères suivants: (1) elle peut être produite dans un certain nombre d’endroits différents, mais elle est accessible en tant qu’entité unique; (2) elle doit être organisée et indexée pour un accès aussi facile que possible à partir du lieu où elle est produite; (3) elle doit être stockée et gérée de manière à avoir une existence assez longue après sa création; (4) elle doit trouver un équilibre entre le respect du droit d’auteur et les exigences universitaires.

Les livres numériques émanant pour la plupart de livres imprimés, ils doivent être numérisés soit en mode image soit en mode texte. La numérisation en mode image consiste à scanner le livre page après page pour offrir un fac-similé numérique de la version imprimée. La présentation originale étant conservée, on peut «feuilleter» le livre à l'écran. La numérisation en mode texte consiste à scanner le livre en mode image, puis à le convertir en mode texte grâce à un logiciel OCR (reconnaissance optique de caractères). Le mode texte ne conserve pas la présentation originale du livre ni celle de la page, mais il permet l'indexation, la recherche textuelle, l'analyse textuelle, une étude comparative entre plusieurs textes ou plusieurs versions du même texte, etc.

Nombre de bibliothèques de toutes tailles constituent leur propre bibliothèque numérique, souvent disponible en accès libre et gratuit sur le web. La Bibliothèque nationale de France (BnF) lance sa bibliothèque numérique Gallica en octobre 1997, avec 3.000 ouvrages du 19e siècle. Gallica offre 90.000 ouvrages et 80.000 images libres de droits en décembre 2006 et un million de documents tous supports confondus en mars 2010.

Google décide de mettre son expertise — et sa force de frappe — au service du livre et lance Google Print en mai 2005 avant de le relancer sous le nom de Google Books (Google Livres) en août 2006. Mais les livres numériques ne sont disponibles que sur le navigateur web de Google et les ayants droit (éditeurs ou auteurs) des livres sous droits ne sont pas sollicités en amont, ce qui donne lieu à quelques procès retentissants avec des associations d'éditeurs, suivis de près par le monde du livre.

En parallèle, d'autres projets «ouverts» voient le jour, comme la bibliothèque numérique Europeana lancée en novembre 2008 par l'Union européenne ou l'Open Content Alliance (Alliance pour un contenu ouvert) lancée en octobre 2006 par l'Internet Archive pour créer une bibliothèque numérique publique à vocation mondiale. Contrairement à Google Livres, ces bibliothèques ne sont pas liées à des enjeux commerciaux, elles sont consultables sur tout moteur de recherche et elles sont respectueuses du droit d'auteur puisqu'elles ne diffusent pas de documents sous droits, à moins que les ayants droit aient dûment donné leur accord.

Europeana, bibliothèque numérique publique européenne, ouvre ses portes en novembre 2008 avec deux millions de documents. Europeana offre 6 millions de documents en mars 2010, 10 millions de documents en septembre 2010 et 39 millions de documents en mai 2015.

L'Internet Archive lance l'Open Content Alliance (OCA) on octobre 2005 afin de fédérer plusieurs organismes partenaires pour créer une bibliothèque numérique gratuite de livres numérisés et de documents multimédias. Pour mémoire, l'Internet Archive est fondée en avril 1996 par Brewster Kahle à San Francisco afin d'archiver le web pour les générations présentes et futures.

Six mois plus tard, l'Open Content Alliance regroupe de nombreux partenaires: des bibliothèques universitaires bien sûr, mais aussi des organisations gouvernementales, des associations à but non lucratif, des organismes culturels et des sociétés informatiques (Adobe, Hewlett Packard, Microsoft, Yahoo!, Xerox, etc.).

Les premiers organismes donnant leur accord pour la numérisation de leurs documents sont les bibliothèques des universités de Californie et de Toronto, les Archives européennes, les Archives nationales du Royaume-Uni, la maison d'édition O'Reilly Media et les Prelinger Archives. Le site de l'Internet Archive offre 100.000 livres numérisés en décembre 2006, avec un rythme de 12.000 nouveaux livres par mois, 200.00 livres numérisés en mai 2007, un million de livres numérisés en décembre 2008 et deux millions de livres numérisés en mars 2010.

Des collections spécialisées sont également numérisées. L'Internet Archive reçoit par exemple une subvention d'un million de dollars US en décembre 2006 de la part de la Sloan Foundation pour numériser les collections du Metropolitan Museum of Art (la totalité des livres et plusieurs milliers d'images) ainsi que plusieurs collections spécifiques de la Bibliothèque publique de Boston (les 3.800 livres de la bibliothèque personnelle de John Adams, deuxième président des États-Unis), du Getty Institute (une collection de livres d'art), de l'Université John Hopkins (une collection de documents sur le mouvement anti-esclavagiste) et de l'Université de Californie à Berkeley (une collection de documents sur la ruée vers l'or).

Les trésors du passé

Les bibliothèques numérisent aussi les trésors du passé pour les offrir au monde, par exemple la Bible de Gutenberg originale, qu'il est possible d'admirer en novembre 2000 sur le site de la British Library. Cette Bible aurait été imprimée en 1454 ou 1455 par Gutenberg en 180 exemplaires dans son atelier de Mayence (Allemagne). 48 exemplaires, dont certains incomplets, existeraient toujours. La British Library possède deux versions complètes et une version partielle de cette Bible. En mars 2000, dix chercheurs et experts techniques de l'Université Keio de Tokyo et de NTT (Nippon Telegraph and Telephone Communications) viennent passer deux semaines sur place pour numériser les deux versions complètes, légèrement différentes, avec mise en ligne quelques mois plus tard.

D'autres trésors de la British Library sont déjà en ligne, par exemple Beowulf, considéré comme le premier poème anglo-saxon, Magna Carta, premier texte constitutionnel anglais signé en 1215, les Lindisfarne Gospels, trésor inestimable datant de 698, le Diamond Sutra, autre trésor inestimable datant de 868, les Sforza Hours, trésor de la Renaissance datant des années 1490-1520, le Codex Arundel, qui regroupe les notes prises par Léonard de Vinci entre 1480 et 1518, ou encore le Tyndale New Testament, qui fut le premier Nouveau Testament en langue anglaise imprimé en 1526.

Un autre trésor mis en ligne est l'Encyclopédie de Diderot, mais cette fois sous la forme d'une base de données créée par le projet ARTFL (American and French Research on the Treasury of the French Language — Recherche américaine et française sur les trésors de la langue française), un projet commun du CNRS (Centre national de la recherche scientifique) en France

et de l'Université de Chicago aux États-Unis. Le projet ARTFL a pour but de constituer une base de données de 2.000 ouvrages des 13e -20e siècles ayant trait à la littérature, à la philosophie, aux arts ou aux sciences.

Mise en ligne en 1998, la base de données du premier volume (1751) de l'Encyclopédie de Diderot permet une recherche par mot, portion de texte et auteur. Des liens permettent de passer des versions en mode texte au fac-similé des pages originales. L'automatisation de la production des versions en mode texte entraîne des erreurs typographiques qui sont corrigées au fil des mois. La recherche d'images est possible aussi dans un deuxième temps.

La mise en ligne expérimentale du premier volume est le prélude à une base de données exhaustive comprenant la première édition (1751-1772) de l 'Encyclopédie dans son entier, à savoir 17 volumes de texte (soit 18.000 pages et 21,7 millions de mots) et 11 volumes de planches, avec des planches d'une telle qualité technique que certaines font toujours référence dans leur domaine au 21e siècle.

L'Encyclopédie de Diderot — dont le nom complet est «Encyclopédie ou Dictionnaire raisonné des sciences, des métiers et des arts de Diderot et d'Alembert» — comprend 72.000 articles rédigés par 140 collaborateurs, dont Rousseau, Voltaire, Marmontel, d'Holbach, Turgot, etc. Une encyclopédie collective donc, bien avant Wikipédia. Monumental ouvrage de référence destiné à rassembler puis divulguer les connaissances de l'époque, l'Encyclopédie porte la marque des courants intellectuels et sociaux du Siècle des lumières. C'est grâce à elle que se propagent les idées nouvelles qui inspireront la Révolution française de 1789.

Diderot explique lui-même dans l'introduction que «le but d'une encyclopédie est de rassembler les connaissances éparses sur la surface de la terre, d'en exposer le système général aux hommes avec qui nous vivons et de le transmettre aux hommes qui viendront après nous, afin (…) que nos neveux, devenant plus instruits, deviennent en même temps plus vertueux et plus heureux, et que ne mourions pas sans avoir bien mérité du genre humain.» Un beau texte qui figure aussi sur un mur de l'Allée de l'Encyclopédie, l'une des grandes artères de la Bibliothèque nationale de France (BnF).

Outre l'Encyclopédie, l'ARTFL propose une base de données exhaustive du «Dictionnaire de l'Académie française», dont les différentes éditions s'échelonnent entre 1694 et 1935. La première édition (1694) et la cinquième édition (1798) du dictionnaire sont les premières à être disponibles en ligne, avec possibilité de recherche par mot puis par portion de texte. Les différentes éditions sont ensuite combinées dans une base de données unique, qui permet de juger de l'évolution d'un terme en consultant aussi bien une édition spécifique que l'ensemble des éditions.

D'autres projets de l'ARTFL concernent par exemple le «Thresor de la langue française» de Jean Nicot (1606), le «Dictionnaire historique et critique» de Philippe Bayle (1740) ou encore un projet biblique multilingue comprenant entre autres «La Bible française» de Louis Segond (1910). Il s'agit là encore de bases de données avec moteur de recherche.

Les catalogues collectifs

L’internet permet non seulement la consultation de catalogues en ligne mais aussi la gestion de catalogues collectifs, qui sont en plein essor à l’échelon local, régional, national ou mondial, avec notices abrégées ou complètes, et recherche simple ou avancée. Le résultat d’une recherche peut être copié, imprimé, sauvegardé et envoyé par courriel.

Le but premier d’un catalogue collectif est d’éviter de cataloguer à nouveau un document déjà traité par une bibliothèque partenaire. Si le catalogueur trouve la notice du livre qu’il est censé cataloguer, il la copie pour l’inclure dans le catalogue de sa propre bibliothèque en y ajoutant les données locales. Si le catalogueur ne trouve pas la notice, il la crée, et cette notice peut être utilisée par les catalogueurs officiant dans d’autres bibliothèques.

Un exemple de catalogue collectif à l’échelle d’un pays est le Catalogue collectif de France (CCFr), mis en chantier en juillet 1997 pour permettre de «trouver des informations détaillées sur les bibliothèques françaises, leurs collections et leurs fonds (anciens, locaux ou spécifiques), connaître précisément les services qu’elles rendent et interroger leur catalogue en ligne». À terme, ce catalogue permettra aussi de «localiser des ouvrages (documents imprimés, audio, vidéo, multimédia) dans les principales bibliothèques et demander le prêt ou la reproduction» de documents qui seront remis à l’usager dans la bibliothèque de son choix. C’est chose faite en novembre 2002.

Le Catalogue collectif de France regroupe les catalogues de la Bibliothèque nationale de France (BnF) et des bibliothèques universitaires françaises, ainsi que les catalogues des fonds anciens (avant 1811) et des fonds locaux des bibliothèques municipales et spécialisées. Sa gestion est confiée à la Bibliothèque nationale de France en juillet 2001. Ce catalogue permet de localiser 15 millions de documents (appartenant à 160 bibliothèques) en décembre 2006 et 30 millions de documents en mai 2015.

L’internet facilite aussi la gestion de catalogues collectifs mondiaux. Deux associations américaines se lancent dans cette tâche titanesque dans les années 1980, l’OCLC (Online Computer Library Center — Centre informatique des bibliothèques en ligne) et le RLG (Research Libraries Group — Groupe des bibliothèques de recherche).

Vingt ans plus tard, l’OCLC et le RLG gèrent de gigantesques bases bibliographiques alimentées par leurs adhérents — et dûment facturées jusqu’à la mise en ligne gratuite de ces catalogues sur le web. RLG lance une version web gratuite de son catalogue au printemps 2004 sous le nom de RedLightGreen, avec 130 millions de notices, deux ans avant le lancement en août 2006 de Worldcat.org, la version web gratuite de WorldCat, la catalogue collectif d’OCLC. Le site RedLightGreen cesse ses activités en novembre 2006 suite à la fusion de RLG avec OCLC, et les usagers de RedLightGreen sont invités à utiliser WorldCat.org.

Worldcat.org permet de localiser 1,5 milliard de documents en avril 2010 et d’avoir directement accès à nombre d’entre eux.

Les ressources linguistiques

Travlang est le premier site web à proposer des dictionnaires bilingues gratuits. Michael C. Martin, alors étudiant en physique, crée d'abord en 1994 la rubrique «Foreign Languages for Travelers» (Langues étrangères pour les voyageurs) sur le site de son université à New York pour inciter d'autres étudiants à apprendre des langues sur le web. Cette rubrique s'étoffe rapidement et rencontre un grand succès.

L'année suivante, Michel C. Martin intègre cette rubrique dans son nouveau site Travlang, dédié à la fois aux voyages et aux langues. Très populaire, Travlang est nommé meilleur site de voyages en 1997. Devenu chercheur en physique au Lawrence Berkeley National Laboratory (Laboratoire national Lawrence à Berkeley) en Californie, Michael C. Martin continue de gérer Travlang lui-même sur son temps libre.

La section «Foreign Languages for Travelers» (Langues étrangères pour les voyageurs) permet d'apprendre les rudiments de 60 langues sur le web en 1998 et la section «Translating Dictionaries» (Dictionnaires de traduction) donne accès à des dictionnaires gratuits dans 15 langues (afrikaans, allemand, danois, espagnol, espéranto, finnois, français, frison, hollandais, hongrois, italien, latin, norvégien, portugais, tchèque). Ces dictionnaires sont le plus souvent sommaires et de qualité inégale.

D'autres sections proposent des liens vers des services de traduction, des écoles de langue, des librairies multilingues, etc. On peut également réserver son hôtel, sa voiture ou son billet d'avion, s'informer sur les taux de change ou encore consulter un répertoire d'autres sites de langues et de voyages.

Michael C. Martin écrit en août 1998 dans un entretien par courriel: «Je pense que le web est un endroit idéal pour rapprocher les cultures et les personnes, et ceci inclut d'être multilingue. Notre site Travlang est très populaire pour cette raison, et les gens aiment être en contact avec d'autres parties du monde. L'internet est un outil très important pour communiquer avec des gens avec lesquels on n'aurait pas l'occasion de dialoguer autrement. J'apprécie vraiment la collaboration générale qui a rendu possibles les pages de “Foreign Languages for Travelers”. Je pense aussi que les traductions intégrales informatisées vont devenir monnaie courante, et qu'elles permettront de communiquer à la base avec davantage de gens. Ceci aidera aussi à amener davantage l'internet vers le monde non anglophone.» Michael C. Martin vend Travlang en février 1999. Le site délaisse sa partie linguistique pour se consacrer uniquement aux voyages.

Tyler Chambers, informaticien à Boston (États-Unis), gère deux projets linguistiques sur son temps libre, la Human-Languages Page (H-LP) depuis 1994 et l'Internet Dictionary Project (IDP) depuis 1995. Il relate en septembre 1998: «Le multilinguisme sur le web était inévitable bien avant que ce médium se développe vraiment. Mon premier vrai contact avec le web date de 1994, un peu après ses débuts mais bien avant son expansion. 1994 a été aussi l'année où j'ai débuté mon premier projet web multilingue [la Human-Languages Page], et il existait déjà

un nombre significatif de ressources linguistiques en ligne, avant même la création de Netscape. Mosaic était le seul navigateur sur le web, et les pages web étaient surtout des documents textuels reliés par des hyperliens.»

La Human-Languages Page (Page des langues humaines) est un répertoire de ressources linguistiques. De qualité très supérieure au répertoire de Travlang, ce répertoire recense 1.800 ressources dans une centaine de langues en octobre 1998. Puis il fusionne au printemps 2001 avec le Languages Catalog (Catalogue des langues), une section de la WWW Virtual Library, pour devenir iLoveLanguages. iLoveLanguages offre 2.000 ressources linguistiques dans une centaine de langues en septembre 2003.

L'Internet Dictionary Project (Projet de dictionnaire internet), le deuxième projet web multilingue de Tyler Chambers, est un projet coopératif ouvert à tous pour créer des dictionnaires bilingues gratuits de l'anglais vers d'autres langues (allemand, espagnol, français, italien, latin, portugais). Comme expliqué sur le site: «Le but est de créer des dictionnaires de traduction grâce à l'aide des usagers. Ce site permet aux usagers du monde entier de consulter ces dictionnaires et de participer à la traduction de termes anglais dans d'autres langues. Les listes de termes anglais et leurs correspondants dans d'autres langues sont ensuite mis à la disposition de tous sur ce site, sans restriction d'aucune sorte.»

«L'Internet Dictionary Project a débuté en 1995 pour combler une lacune et procurer des dictionnaires de traduction gratuits à la communauté internet et à tous ceux qui s'intéressent à l'informatique. Non seulement il est très utile d'avoir immédiatement accès à des dictionnaires via le World Wide Web, mais cela permet aussi le développement de logiciels pouvant tirer parti de tels dictionnaires, que ce soit des programmes de traduction ou des vérificateurs d'orthographe ou encore des guides d'apprentissage des langues. En facilitant la création de ces dictionnaires en ligne par des milliers de volontaires, et en les mettant gratuitement à la disposition de tous, l'Internet Dictionary Project espère imprimer sa marque sur l'internet et susciter d'autres projets qui seront plus bénéfiques que de générer des revenus purement financiers.»

Tyler Chambers raconte en septembre 1998 dans un entretien par courriel: «Bien que je ne sois pas multilingue, ni même bilingue moi-même, je suis conscient du fait que très peu de domaines ont une importance comparable à celle des langues et du multilinguisme. Je pense que le web est important pour la sensibilisation aux langues et pour les questions culturelles. Dans quel autre endroit peut-on chercher au hasard pendant vingt minutes et trouver des informations susceptibles de vous intéresser dans trois langues différentes sinon plus? Dire que l'internet favorise le multilinguisme est à mon sens une opinion fausse. C'est la communication qui favorise le multilinguisme et l'échange multiculturel. L'internet est seulement le mode de communication le plus récent qui soit accessible au grand public. Les langues deviendront encore plus importantes lorsque tout le monde pourra communiquer à l'échelle de la planète (à travers le web, les discussions, les jeux, le courrier électronique, ou toute application appartenant encore au domaine de l'avenir). Avec l'amélioration des navigateurs web et l'expérience acquise par les usagers, je ne pense pas qu'il existe une langue vivante qui ne soit pas maintenant représentée sur le web, que ce soit la langue des Indiens d'Amérique ou les dialectes moyen-orientaux. De même, une pléthore de langues mortes peut maintenant trouver une audience nouvelle avec des érudits et autres spécialistes en ligne.»

Tyler Chambers met fin à l'Internet Dictionary Project en janvier 2007, faute de temps, tout en laissant les dictionnaires existants tels quels pour consultation ou téléchargement.

La grande société de traduction Logos décide fin 1997 de mettre ses outils professionnels en accès libre sur le web pour tous ses traducteurs et pour le grand public. Ces outils professionnels sont: (1) le Logos Dictionary (Dictionnaire Logos), un dictionnaire multilingue de 7,5 millions de termes (en septembre 1998); (2) la Wordtheque, une base de données multilingue de 328 millions de termes, constituée à partir de milliers de traductions (romans, documents techniques et autres), avec une recherche possible par langue, mot, auteur ou titre; (c) Linguistic Resources (Ressources linguistiques), un point d'accès unique à 553 glossaires; (3) l'Universal Conjugator (Conjugueur universel), avec des tableaux de conjugaison dans 17 langues.

Fondé en 1979 par Rodrigo Vergara à Modène (Italie), Logos propose des services de traduction dans 35 langues en 1997, avec 300 traducteurs travaillant sur place et un réseau mondial de 2.500 traducteurs travaillant en free-lance. La moyenne de production est de 200 textes par jour.

Interviewé par la journaliste Annie Kahn pour son article «Les mots pour le dire» paru dans le quotidien Le Monde du 7 décembre 1997, Rodrigo Vergara relate: «Nous voulions que nos traducteurs aient tous accès aux mêmes outils de traduction. Nous les avons donc mis à leur disposition sur internet, et tant qu'à faire nous avons ouvert le site au public. Cela nous a rendus très populaires, nous a fait beaucoup de publicité. L'opération a drainé vers nous de nombreux clients, et nous a permis aussi d'étoffer notre réseau de traducteurs grâce aux contacts établis à la suite de cette initiative.»

Annie Kahn explique aussi dans le même article: «Le site de Logos est beaucoup plus qu'un dictionnaire ou qu'un répertoire de liens vers d'autres dictionnaires en ligne. L'un des piliers du système est un logiciel de recherche documentaire fonctionnant sur un corpus de textes littéraires disponibles gratuitement sur internet. Lorsqu'on recherche la définition ou la traduction d'un mot, "didactique" par exemple, on trouve non seulement le résultat recherché, mais aussi une phrase d'une oeuvre littéraire utilisant ce mot (en l'occurrence, un essai de Voltaire). Un simple clic permet d'accéder au texte intégral de l'oeuvre ou de commander le livre grâce à un partenariat avec Amazon.com, le libraire en ligne bien connu. Il en est de même avec les traductions étrangères. Si aucun texte utilisant ce mot n'a été trouvé, le système fonctionne alors comme un moteur de recherche et renvoie aux sites web concernant ce mot. Pour certains termes, il est proposé d'en entendre la prononciation. Si une traduction manque, le système fait un appel au peuple. À chacun d'enrichir la base, les traducteurs de l'entreprise valident ensuite les traductions proposées.»

Dix ans plus tard, en 2007, la Wordtheque, devenue la Logos Library (Bibliothèque de Logos), comprend 710 millions de termes. Linguistic Resources, qui n'a pas changé de nom, offre un point d'accès unique à 1.215 glossaires. L'Universal Conjugator, devenu Conjugation of Verbs (Conjugaison de verbes), propose des tableaux de conjugaison dans 36 langues.

Les dictionnaires

Après les premiers dictionnaires en ligne bricolés avec peu de moyens, aussi bien pour leur contenu que pour leur présentation, on voit apparaître des dictionnaires de renom, le plus souvent issus d’ouvrages imprimés.

Mis en ligne dès 1997 avec accès libre et gratuit, le Dictionnaire universel francophone en ligne répertorie 45.000 mots et 116.000 définitions tout en présentant «sur un pied d’égalité, le français dit “standard” et les mots et expressions en français tel qu’on le parle sur les cinq continents». Issu de la collaboration entre Hachette et l’Agence universitaire de la francophonie (AUF), il correspond à la partie «noms communs» du dictionnaire imprimé publié par Hachette.

L’équivalent pour la langue anglaise est le site Merriam-Webster OnLine, qui donne librement accès aux versions en ligne du Collegiate Dictionary (Dictionnaire collégial) et du Collegiate Thesaurus (Thésaurus collégial), deux ouvrages imprimés de référence publiés par Merriam-Webster.

L’Oxford University Press (Presses de l’Université d’Oxford) met en ligne en mars 2000 l’équivalent numérique des 20 volumes de l’Oxford English Dictionary (Dictionnaire anglais d’Oxford), avec une consultation payante qui inclut la mise à jour trimestrielle de mille entrées nouvelles ou révisées.

Conçu directement pour le web, avec accès libre et gratuit, le Grand dictionnaire terminologique (GDT) est une initiative de l’Office québécois de la langue française (OQLF) visant à proposer un dictionnaire bilingue français-anglais de haut niveau. Mis en ligne en septembre 2000, le GDT est précédé deux ans plus tôt par Le Signet, une base terminologique des technologies de l’information, dont les 10.000 fiches bilingues sont ensuite intégrées au GDT. Le GDT propose d’emblée 3 millions de termes appartenant au vocabulaire industriel, scientifique et commercial. Sa mise en ligne est le résultat d’un partenariat entre l’OQLF, auteur du dictionnaire, et Semantix, société spécialisée dans les solutions logicielles linguistiques. Événement célébré par de nombreux linguistes, cette mise en ligne est un succès. Dès le premier mois, le GDT est consulté par 1,3 million de personnes, avec 60.000 requêtes par jour. La gestion du dictionnaire est ensuite assurée par Convera Canada, avec 3,5 millions de requêtes mensuelles en février 2003. Une nouvelle version du GDT est mise en ligne en mars 2003, avec une gestion désormais assurée par l’OQLF lui-même, et l’ajout du latin comme troisième langue.

Professeur de langues à l’Université Bucknell (États-Unis), Robert Beard co-fonde en février 2000 le portail yourDictionary.com, qui intègre son premier site, A Web of Online Dictionaries, créé dès 1995 pour offrir un répertoire de liens vers des dictionnaires de langues en ligne et d’autres outils linguistiques (dictionnaires multilingues, dictionnaires anglophones spécialisés, thésaurus, vocabulaires, grammaires, glossaires, méthodes de langues), sans oublier la section Linguistic Fun (Plaisir linguistique), réservée aux non spécialistes.

Robert Beard raconte en septembre 1998 dans un entretien par courriel: «On a d’abord craint que le web représente un danger pour le multilinguisme, étant donné que le langage HTML et d’autres langages de programmation sont basés sur l’anglais et qu’on trouve tout simplement davantage de sites web en anglais que dans toute autre langue. Cependant, le site web que je gère montre que le multilinguisme est très présent et que le web peut en fait permettre de préserver des langues menacées de disparition. Je propose maintenant des liens vers des dictionnaires dans 150 langues différentes et des grammaires dans 65 langues différentes. De plus, comme les concepteurs de logiciels de navigation manifestent une attention nouvelle pour la diversité des langues dans le monde, ceci va encourager la présence de davantage encore de sites web dans différentes langues.»

Robert Beard relate aussi en janvier 2000 lors du lancement du portail yourDictionary.com: «Nos nouvelles idées sont nombreuses. Nous projetons de travailler avec l’Endangered Language Fund [Fonds pour les langues en danger] aux États-Unis et en Grande-Bretagne afin de rassembler des fonds pour cette fondation et nous publierons les résultats sur notre site. Nous aurons des groupes de discussion et des bulletins d’information sur les langues. Il y aura des jeux de langue destinés à se distraire et à apprendre les bases de la linguistique. La page Linguistic Fun deviendra une lettre d’information en ligne avec des extraits courts, intéressants et même amusants dans différentes langues, choisis par des experts du monde entier.»

Soucieux de servir toutes les langues sans exception, le portail propose la section Endangered Language Repository [Répertoire des langues en danger], consacrée aux langues menacées. Robert Beard écrit à la même date: «Les langues menacées sont essentiellement des langues non écrites. Un tiers seulement des quelque 6.000 langues existant dans le monde sont à la fois écrites et parlées. Je ne pense pourtant pas que le web va contribuer à la perte de l’identité de ces langues et j’ai même le sentiment que, à long terme, il va renforcer cette identité. Par exemple, de plus en plus d’Indiens d’Amérique contactent des linguistes pour leur demander d’écrire la grammaire de leur langue et de les aider à élaborer des dictionnaires. Pour eux, le web est un instrument à la fois accessible et précieux d’expression culturelle.»

yourDictionary.com répertorie 1.800 dictionnaires dans 250 langues en septembre 2003, ainsi que de nombreux outils linguistiques (vocabulaires, grammaires, glossaires, méthodes de langues, etc.). Le portail propose 2.500 dictionnaires et grammaires dans 300 langues en avril 2007.

Michael Kellogg crée le site WordReference.com en 1999 pour proposer des dictionnaires bilingues gratuits en ligne. Il raconte sur le site: «J’ai débuté ce site en 1999 pour offrir des dictionnaires bilingues gratuits en ligne et d’autres outils pour tous sur l’internet. Depuis, le site s’est progressivement développé pour devenir l’un des sites de dictionnaires en ligne les plus utilisés, et le principal dictionnaire en ligne pour les paires de langues anglais-espagnol, anglais-français, anglais-italien, espagnol-français et espagnol-portugais. Mon site est toujours classé sans interruption parmi les 500 sites les plus visités du web. Aujourd’hui, je suis heureux de continuer à améliorer ces dictionnaires, les autres outils linguistiques du site et les forums de langues. J’ai vraiment plaisir à créer de nouvelles fonctionnalités pour rendre ce site de plus en plus utile.»

Les dictionnaires les plus populaires sont les dictionnaires espagnol-anglais, français-anglais et italien-anglais. Pour l'anglais, WordReference.com propose un dictionnaire monolingue et des dictionnaires de l'anglais vers d'autres langues (arabe, chinois, coréen, français, grec, italien, japonais, polonais, portugais, roumain, tchèque, turc) et vice versa. Pour l'espagnol, on trouve un dictionnaire monolingue, un dictionnaire de synonymes, un dictionnaire espagnol-français et un dictionnaire espagnol-portugais. On trouve aussi des dictionnaires monolingues pour l'allemand et le russe. Des tableaux de conjugaison sont disponibles pour l'espagnol, le français et l'italien.

Lancée en 2010, WordReference Mini est une version miniature du site qui permet son intégration dans d'autres sites, par exemple des sites d'apprentissage de langues. Une version pour appareil mobile existe aussi pour les dictionnaires les plus consultés (anglais-espagnol, anglais-français, anglais-italien et vice versa), avec d'autres paires de langues à venir.

Les encyclopédies

On voit apparaître fin 1999 des encyclopédies en ligne de renom, le plus souvent issues de leurs équivalents imprimés, avant la création d'encyclopédies directement conçues pour le web.

WebEncycIo, publié par les éditions Atlas, est la première grande encyclopédie francophone en accès libre, avec mise en ligne en décembre 1999. La recherche est possible par mot-clé, par thème et par média (carte, lien internet, photo, illustration). Un appel à contribution incite les spécialistes d'un sujet donné à envoyer des articles, qui sont ensuite publiés dans la section WebEncycIo contributif. Après avoir été libre, l'accès est ensuite soumis à une inscription préalable gratuite.

La version web de l'Encyclopaedia Universalis est elle aussi mise en ligne en décembre 1999, avec 28.000 articles signés par 4.000 auteurs. Si la consultation est payante sur la base d'un abonnement annuel, de nombreux articles sont en accès libre.

Mis en ligne à la même date, le site Britannica.com propose l'équivalent numérique des 32 volumes de l'Encyclopaedia Britannica (15e édition) en accès libre et gratuit, en complément de la version imprimée et de la version CD-ROM, toutes deux payantes. Le site offre également une sélection d'articles issus de 70 magazines, un guide des meilleurs sites, un choix de livres, etc., le tout étant accessible à partir d'un moteur de recherche commun. En septembre 2000, le site fait partie des cent sites les plus visités du web. En juillet 2001, la consultation devient payante sur la base d'un abonnement mensuel ou annuel. Beaucoup plus tard, en 2009, Britannica.com ouvre son site à des contributeurs externes, avec inscription obligatoire pour écrire ou modifier des articles.

Deux ans après la mise en ligne de l'Oxford English Dictionary (Dictionnaire anglais d'Oxford), l'Oxford University Press (Presses de l'Université d'Oxford) lance en mars 2002 l'Oxford Reference Online (Référence en ligne d'Oxford), une vaste encyclopédie conçue

directement pour le web et consultable sur abonnement payant. Avec 60.000 pages et un million d'entrées, elle représente l'équivalent d'une centaine d'ouvrages de référence imprimés.

Arrivent ensuite les encyclopédies en ligne participatives, avec en tête de file Wikipédia, fondée en janvier 2001 par Jimmy Wales et Larry Sanger (Larry Sanger quitte plus tard l'équipe), qui devient rapidement l'un des sites les plus visités du web.

Wikipédia est une encyclopédie gratuite en ligne écrite collectivement et dont le contenu est librement réutilisable. Sans publicité et financée par des dons, elle est rédigée par des milliers de volontaires, avec possibilité d'écrire, de corriger et de compléter les articles, aussi bien les siens que ceux d'autres contributeurs. Les articles restent la propriété de leurs auteurs et leur libre utilisation est régie par une licence GFDL puis par une licence Creative Commons.

En décembre 2004, Wikipédia compte 1,3 million d'articles rédigés par 13.000 contributeurs dans une centaine de langues. En décembre 2006, Wikipédia compte 6 millions d'articles dans 250 langues et devient l'un de dix sites les plus visités du web. En mai 2007, 7 millions d'articles sont disponibles dans 192 langues, dont 1,8 million d'articles en anglais, 589.000 articles en allemand, 500.000 articles en français, 260.000 articles en portugais et 236.000 articles en espagnol. En 2008, Wikipédia est l'un des cinq sites les plus visités du web. En septembre 2010, Wikipédia compte 14 millions d'articles dans 272 langues, dont 3,4 millions d'articles en anglais, 1,1 million d'articles en allemand et 1 million d'articles en français. Wikipédia fête ses dix ans en janvier 2011 avec 17 millions d'articles dans 270 langues et 400 millions de visiteurs par mois pour l'ensemble de ses sites.

Créée en juin 2003, la Wikimedia Foundation gère non seulement Wikipédia mais aussi Wiktionary (dictionnaire et thésaurus multilingue), Wikibooks (livres et manuels), Wikiquote (répertoire de citations), Wikisource (textes du domaine public), Wikimedia Commons (sources multimédia), Wikispecies (répertoire d'espèces animales et végétales), Wikinews (site d'actualités) et Wikiversity (matériel d'enseignement).

Wikipédia inspire aussi bien d'autres projets au fil des ans, par exemple Citizendium, créé en mars 2007 par Larry Sanger en tant qu'encyclopédie collaborative expérimentale au contenu vérifié par des experts, ou encore l'Encyclopedia of Life (Encyclopédie de la vie), créée en mai 2007 pour recenser toutes les espèces animales et végétales connues.

Dans «Why make room for experts in web 2.0?» (Pourquoi accorder une place aux experts dans le web 2.0?), un essai en ligne daté d'octobre 2006 et actualisé en mars 2007, Larry Sanger voit dans Citizendium l'émergence d'un nouveau modèle de collaboration massive de dizaines de milliers de personnes, non seulement pour les encyclopédies, mais aussi pour les manuels d'enseignement, les ouvrages de référence, les projets multimédias et les applications 3D. Cette collaboration est basée sur le partage des connaissances, dans la lignée du web 2.0, mais avec la vérification de ces connaissances par des experts. D'après Larry Sanger, des structures de ce type pourraient être également créées pour des collaborations scientifiques et médicales, et Citizendium pourrait servir de prototype dans ce domaine.

Lancée en mai 2007, avec une version pilote disponible sur le web un an plus tard, l'Encyclopedia of Life est une encyclopédie collaborative en ligne conçue pour rassembler les connaissances sur toutes les espèces animales et végétales connues (1,8 million), y compris les

espèces en voie d’extinction, avec l’ajout de nouvelles espèces au fur et à mesure de leur identification. Cette encyclopédie multimédia vise à rassembler textes, photos, cartes, bandes sonores et vidéos, avec une page web par espèce, afin d’offrir un portail unique à des millions de documents épars, en ligne et hors ligne.

L’Encyclopedia of Life est à destination de tous: scientifiques, enseignants, étudiants, scolaires, médias, décideurs et grand public. Son but est d’être un macroscope permettant de déceler les grandes tendances à partir d’un stock d’informations considérable, à la différence du microscope permettant l’étude de détail. La version initiale en anglais devrait être traduite dans plusieurs langues par des organismes partenaires.

Les revues scientifiques

Entre 2003 et 2007, la Public Library of Science (PLOS — Bibliothèque publique des sciences) fonde plusieurs revues scientifiques et médicales gratuites de haut niveau: PLOS Biology, PLOS Medicine, PLOS Genetics, PLOS Computational Biology, PLOS Pathogens, PLOS Neglected Tropical Diseases (maladies tropicales négligées) et PLOS ONE.

Ces différents titres ne tardent pas à rivaliser avec les meilleures revues scientifiques payantes (et hors de prix), avec l’avantage énorme d’être accessibles à tous. De plus, tous les articles de PLOS utilisent la licence Creative Commons CC BY, c’est-à-dire la plus large qui soit. Ces articles peuvent être librement diffusés et réutilisés ailleurs, y compris pour des traductions, la seule contrainte étant la mention des auteurs, du titre et de la source.

La Public Library of Science est l’un des nombreux organismes qui participent au vaste mouvement mondial promouvant l’accès ouvert à la recherche.

Signée en février 2002, l’Initiative de Budapest pour l’accès ouvert (BOAI — Budapest Open Access Initiative) définit l’accès ouvert ainsi: «Il existe de nombreux degrés et de nombreuses formes d’accès plus large et plus facile à la littérature scientifique. Par “accès ouvert” à cette littérature, nous entendons sa libre mise à disposition sur l’internet public, permettant à tout usager de lire, télécharger, copier, diffuser et imprimer ces articles, de lancer une recherche dans ces articles, de créer un lien vers le texte intégral de ces articles, de les compiler pour les indexer, de les convertir en données pour traitement logiciel, et de les utiliser à toute autre fin légale, sans barrières financières, juridiques ou techniques autres que celles de l’accès à l’internet lui-même. La seule contrainte de reproduction et de diffusion de cette littérature et le seul rôle du droit d’auteur dans ce domaine devraient être de donner aux auteurs le moyen de contrôler l’intégrité de leur travail et le droit d’être mentionnés et cités de manière adéquate» (traduction personnelle).

Créée en 2003, la Directory of Open Access Journals (Répertoire des périodiques en accès ouvert) est un répertoire mondial en ligne de revues scientifiques en accès ouvert dans de nombreuses langues.

Outre les revues ouvertes, de plus en plus d'universités et de centres de recherche ont leur propre archive ouverte, qui leur permet de mettre les publications de leurs professeurs et chercheurs à la disposition de tous. C'est le cas d'universités de réputation mondiale comme Harvard avec DASH (Digital Access to Scholarship at Harvard — Accès numérique au savoir à Harvard) ou le Massachusetts Institute of Technology (Institut de technologie du Massachusetts) avec DSpace@MIT.

Les archives ouvertes sont soit institutionnelles soit thématiques, par exemple arXiv pour la physique ou PubMed pour la biomédecine. Certaines archives ouvertes ne sont pas limitées aux articles et peuvent inclure aussi des thèses, des mémoires, des livres, du matériel pédagogique et des fichiers audio et vidéo, entre autres. Le nombre d'archives ouvertes est en augmentation constante, avec deux grands répertoires mondiaux, ROAR (Registry of Open Access Repositories — Registre des dépôts en accès ouvert) et OpenDOAR (Directory of Open Access Repositories — Annuaire des dépôts en accès ouvert).

Les ressources pour enseigner

Vinton Cerf, co-créateur des protocoles de l'internet en 1974 puis directeur de l'Internet Society (ISOC) en 1992, explique en janvier 1998 lors d'un entretien avec le quotidien Libération: «Le réseau fait deux choses (…): comme les livres, il permet d'accumuler de la connaissance. Mais, surtout, il la présente sous une forme qui la met en relation avec d'autres informations alors que, dans un livre, l'information est maintenue isolée.»

Lors d'une conférence organisée en septembre 1996 par la Fédération internationale du traitement de l'information (IFIP), Dale Spender, professeure et chercheuse australienne, tente de cerner les changements apportés par l'internet dans l'acquisition du savoir et les méthodes d'enseignement. Voici l'argumentaire de sa communication «Creativity and the Computer Education Industry» (La créativité dans l'enseignement par l'informatique) résumé en deux paragraphes.

Pendant plus de cinq siècles, l'enseignement est principalement basé sur l'information donnée par les livres. Or les habitudes liées à l'imprimé ne peuvent être transférées au monde numérique. L'enseignement en ligne offre des possibilités telles qu'il n'est guère possible d'opérer les distinctions traditionnelles entre enseignant et enseigné. Le passage de la culture imprimée à la culture numérique exige d'entièrement repenser le processus d'enseignement, puisque nous avons maintenant l'opportunité sans précédent de pouvoir influer sur le type d'enseignement que nous souhaitons.

Dans la culture imprimée, l'information contenue dans les livres restait la même pendant un certain temps, ce qui nous a encouragés à penser que l'information était stable. La nature même de l'imprimé est liée à la notion de vérité, stable elle aussi. Cette stabilité et l'ordre qu'elle engendre sont l'un des fondements de l'âge industriel et de la révolution scientifique. Les notions de vérité, de loi, d'objectivité et de preuve sont les éléments de référence de nos

croyances et de nos cultures. Mais la révolution numérique change tout ceci. Soudain l'information en ligne supplante l'information imprimée pour devenir la plus fiable et la plus utile, et l'usager est prêt à la payer en conséquence. C'est cette transformation radicale dans la nature de l'information qui doit être au coeur du débat relatif aux méthodes d'enseignement.

En témoigne l'expérience de Patrick Rebollar, professeur de français et de littérature française à Tokyo. Il utilise l'ordinateur pour ses activités d'enseignement et de recherche dès 1987. Il voit apparaître l'internet «dans le champ culturel et linguistique francophone» en 1994 et il débute un site web de recherches et activités littéraires en 1996. Il raconte en juillet 1998 dans un entretien par courriel: «Mon travail de recherche est différent, mon travail d'enseignant est différent, mon image en tant qu'enseignant-chercheur de langue et de littérature est totalement liée à l'ordinateur, ce qui a ses bons et ses mauvais côtés (surtout vers le haut de la hiérarchie universitaire, plutôt constituée de gens âgés et technologiquement récalcitrants). J'ai cessé de m'intéresser à certains collègues proches géographiquement mais qui n'ont rien de commun avec mes idées, pour entrer en contact avec des personnes inconnues et réparties dans différents pays (et que je rencontre parfois, à Paris ou à Tokyo, selon les vacances ou les colloques des uns ou des autres). La différence est d'abord un gain de temps pour tout, puis un changement de méthode de documentation, puis un changement de méthode d'enseignement privilégiant l'acquisition des méthodes de recherche par mes étudiants, au détriment des contenus (mais cela dépend des cours). Progressivement, le paradigme réticulaire l'emporte sur le paradigme hiérarchique.»

Robert Beard, professeur à l'Université Bucknell (États-Unis) et créateur de sites linguistiques (A Web of Online Dictionaries et yourDictionary.com), écrit en septembre 1998: «En tant que professeur de langue, je pense que le web présente une pléthore de nouvelles ressources disponibles dans la langue étudiée, de nouveaux instruments d'apprentissage (exercices interactifs Java et Shockwave) et de test, qui sont à la disposition des étudiants lorsque ceux-ci en ont le temps ou l'envie, 24 heures/24 et 7 jours/7. Aussi bien pour mes collègues que pour moi, et bien sûr pour notre établissement, l'internet nous permet aussi de publier pratiquement sans limitation. L'internet nous offrira tout le matériel pédagogique dont nous pouvons rêver, y compris des notes de lecture, exercices, tests, évaluations et exercices interactifs plus efficaces que par le passé parce que reposant davantage sur la notion de communication. Le web sera une encyclopédie du monde faite par le monde pour le monde. Il n'y aura plus d'informations ni de connaissances utiles qui ne soient pas disponibles, si bien que l'obstacle principal à la compréhension internationale et interpersonnelle et au développement personnel et institutionnel sera levé. Il faudrait une imagination plus débordante que la mienne pour prédire l'effet de ce développement sur l'humanité.»

Un service spécifique de l'Institut des langues de l'Université de Hull (Royaume-Uni) vise à promouvoir l'utilisation des ordinateurs dans l'enseignement des langues. Ce service donne des informations sur la manière dont l'apprentissage des langues assisté par ordinateur peut être intégré aux cours existants et il offre un soutien aux professeurs qui utilisent — ou souhaitent utiliser — l'informatique dans leurs cours. Selon June Thompson, responsable du service, interviewé en décembre 1998: «L'utilisation de l'internet a apporté une nouvelle dimension à notre tâche, qui consiste à aider les professeurs de langue à utiliser les nouvelles technologies dans ce domaine. Avec l'internet, on a la possibilité de davantage utiliser les langues étrangères.

À mon avis, dans un avenir proche, l'utilisation de l'internet pour les langues va continuer à se développer en même temps que d'autres supports dans le cadre d'activités à caractère pédagogique.»

Professeur au département des études françaises de l'Université de Toronto (Canada), Russon Wooldridge explique en février 2001: «Mes activités de recherche, autrefois menées dans une tour d'ivoire, se font maintenant presque uniquement par des collaborations locales ou à distance. Tout mon enseignement exploite au maximum les ressources d'internet (le web et le courriel): les deux lieux communs d'un cours sont la salle de classe et le site du cours, sur lequel je mets tous les matériaux des cours. Je mets toutes les données de mes recherches des vingt dernières années sur le web (réédition de livres, articles, textes intégraux de dictionnaires anciens en bases de données interactives, de traités du 16e siècle, etc.). Je publie des actes de colloques, j'édite un journal, je collabore avec des collègues français, mettant en ligne à Toronto ce qu'ils ne peuvent pas publier en ligne chez eux. Je me rends compte que sans internet mes activités seraient bien moindres, ou du moins très différentes de ce qu'elles sont actuellement. Donc je ne vois pas l'avenir sans.»

Mise en ligne en septembre 2002, la version pilote du MIT OpenCourseWare offre en accès libre le matériel pédagogique de 32 cours représentatifs des cinq facultés du MIT (Massachusetts Institute of Technology — Institut de technologie du Massachusetts). Ce matériel pédagogique comprend des textes de conférences, des travaux pratiques, des exercices et corrigés, des bibliographies, des documents audio et vidéo, etc. Le lancement officiel de l'OpenCourseWare a lieu un an plus tard, avec accès au matériel pédagogique d'une centaine de cours en septembre 2003, 500 cours en mars 2004, 1.400 cours en mai 2006 et 1.800 cours en novembre 2007. Tous ces cours sont régulièrement actualisés, et certains cours sont traduits en espagnol, en portugais et en chinois avec l'aide d'autres organismes.

Le MIT espère que cette expérience — la première du genre — va inciter d'autres universités à promouvoir la mise à disposition gratuite du matériel pédagogique de leurs propres cours. À cet effet, le MIT lance en décembre 2005 l'OpenCourseWare Consortium, avec accès libre et gratuit au matériel d'enseignement de cent universités dans le monde un an plus tard. Renommé Open Education Consortium (Consortium pour un enseignement ouvert), ce consortium comprend 280 organismes participants dans 40 pays en mai 2015, avec 30.000 modules d'enseignement dans 29 langues.

Les ressources pour les traducteurs

Marcel Grangier, directeur de la section française des Services linguistiques centraux de l'Administration fédérale suisse, explique en janvier 1999 dans un entretien par courriel: «Travailler sans internet est devenu tout simplement impossible. Au-delà de tous les outils et commodités utilisés (messagerie électronique, consultation de la presse électronique, activités de services au profit de la profession des traducteurs), internet reste pour nous une source indispensable et inépuisable d'informations dans ce que j'appellerais le "secteur non structuré"

de la toile. Pour illustrer le propos, lorsque aucun site comportant de l’information organisée ne fournit de réponse à un problème de traduction, les moteurs de recherche permettent dans la plus grande partie des cas de retrouver le chaînon manquant quelque part sur le réseau.»

L’équipe de traducteurs francophones gère entre autres Dictionnaires électroniques, une liste quasi exhaustive de dictionnaires monolingues (allemand, anglais, espagnol, français, italien), bilingues et multilingues disponibles en ligne, complétée par des répertoires d’abréviations et acronymes et des répertoires géographiques, essentiellement des atlas. D’abord à usage interne, ce répertoire est disponible ensuite en accès libre sur le web.

Marcel Grangier précise en janvier 2000: «Les Dictionnaires électroniques ne sont qu’une partie de notre site web, et d’autres secteurs ont trait à l’administration, au droit, à la langue française, etc., sans parler des informations générales. Conçu d’abord comme un service intranet, notre site web se veut en premier lieu au service des traducteurs opérant en Suisse, qui souvent travaillent sur la même matière que les traducteurs de l’Administration fédérale, mais également, par certaines rubriques, au service de n’importe quel autre traducteur où qu’il se trouve.» La rubrique Dictionnaires électroniques déménage en 2001 sur le nouveau site de la Conférence des services de traduction des États européens (CST).

Maria Victoria Marinetti, de nationalité mexicaine, est professeure d’espagnol et traductrice. Elle écrit en août 1999: «J’ai accès à un nombre important d’informations au niveau mondial, ce qui est très intéressant pour moi. J’ai également la possibilité de transmettre ou de recevoir des fichiers, des lettres, des photos, etc., dans un va-et-vient d’information constant. L’internet me permet de recevoir ou d’envoyer des traductions générales ou techniques du français vers l’espagnol et vice versa, ainsi que des textes espagnols à reviser. Dans le domaine technique ou chimique, je propose une aide technique, ainsi que des informations sur l’exportation d’équipes de haute technologie vers le Mexique ou d’autres pays d’Amérique latine.»

Praetorius, société londonienne de traduction et services linguistiques, lance le magazine Language Today (Les langues aujourd’hui) à destination des linguistes (traducteurs, interprètes, terminologues, lexicographes, rédacteurs techniques), avec une version imprimée et une version en ligne. Geoffrey Kingscott, directeur de Praetorius, explique en septembre 1998: «Nous publions la version imprimée de Language Today uniquement en anglais, dénominateur commun de nos lecteurs. Quand nous utilisons un article qui était originellement dans une autre langue que l’anglais, ou que nous relatons un entretien conduit dans une autre langue que l’anglais, nous le traduisons en anglais et nous ne publions que la version anglaise, pour la raison suivante: le nombre de pages que nous pouvons imprimer est limité, et déterminé en fonction de notre clientèle (annonceurs et abonnés). Par contre, dans notre version web, nous proposons aussi la version originale.»

Créée dès 1990 par deux professeurs de linguistique, Anthony Rodrigues Aristar et Helen Dry, la principale liste de diffusion pour les linguistes est la Linguist List. La liste débute son propre site web en 1997. Helen Dry explique en août 1998: «La Linguist List, que je modère, a pour politique d’accepter des informations dans toutes les langues, puisque c’est une liste pour linguistes. Nous ne souhaitons cependant pas que le message soit publié dans plusieurs langues, tout simplement à cause de la surcharge de travail que cela représenterait pour notre personnel de rédaction. Nous ne sommes pas une liste fourre-tout, mais une liste modérée. Avant d’être

publié, chaque message est classé par nos étudiants-rédacteurs dans une section comprenant des messages du même type. Notre expérience nous montre que pratiquement tout le monde choisit de publier en anglais. Mais nous relions ces informations à un service de traduction automatique qui présente nos pages dans cinq langues différentes. Ainsi un abonné ne lit la Linguist List en anglais que s'il le souhaite. Nous essayons aussi d'avoir au moins un étudiant-éditeur qui soit réellement multilingue, afin que les usagers puissent correspondre avec nous dans d'autres langues que l'anglais.»

Helen Dry ajoute en juillet 1999: «Nous commençons maintenant à rassembler un grand nombre de données linguistiques. Nous gérons plusieurs bases de données avec moteur de recherche: résumés de thèses de linguistique, informations sur les programmes universitaires de linguistique, informations professionnelles sur les linguistes, etc. À ma connaissance, le fichier des résumés de thèses est la seule compilation électronique qui soit disponible gratuitement sur l'internet.»

Les bases terminologiques

NetGlos (Multilingual Glossary of Internet Terminology — Glossaire multilingue de la terminologie de l'internet) est un glossaire coopératif lancé en 1995 à l'initiative du WorldWide Language Institute (Institut des langues du monde entier). Trois ans plus tard, outre l'anglais, NetGlos est disponible dans douze langues (allemand, chinois, croate, espagnol, français, grec, hébreu, hollandais, italien, maori, norvégien, portugais), avec la participation de nombreux traducteurs de par le monde.

Brian King, directeur de l'Institut, explique en septembre 1998 dans un entretien par courriel: «Une grande partie de la terminologie technique disponible sur le web n'est pas encore traduite dans d'autres langues [que l'anglais]. Et, comme nous nous en sommes rendus compte dans NetGlos, la traduction de ces termes n'est pas toujours facile. Avant qu'un nouveau terme soit accepté comme le terme correct, il y a une période d'instabilité avec plusieurs candidats en compétition. Souvent un terme emprunté à l'anglais est le point de départ et, dans de nombreux cas, il est aussi le point d'arrivée. Finalement émerge un vainqueur qui est ensuite utilisé aussi bien dans les dictionnaires techniques que dans le vocabulaire quotidien de l'usager non spécialiste. La dernière version de NetGlos est la version russe, et elle devrait être disponible dans deux semaines environ. Elle sera sans nul doute un excellent exemple du processus dynamique en cours pour la russification de la terminologie du web.»

«Notre projet NetGlos dépend du bon vouloir de traducteurs bénévoles dans de nombreux pays: Canada, États-Unis, Autriche, Norvège, Belgique, Israël, Portugal, Russie, Grèce, Brésil, Nouvelle-Zélande, etc. À mon avis, les centaines de visiteurs qui consultent quotidiennement les pages de NetGlos constituent un excellent témoignage du succès de ce type de relations de travail. La coopération s'accroîtra encore plus à l'avenir, mais pas nécessairement sur la base du volontariat. La technologie change à une allure frénétique. L'apprentissage durant toute la vie est une stratégie que nous devons tous adopter si nous voulons rester en tête et être compétitifs.

C’est une tâche qui est déjà assez difficile dans un environnement anglophone. Si nous ajoutons à cela la complexité de communiquer dans un cyberespace multilingue et multiculturel, la tâche devient encore plus astreignante. Probablement davantage encore que par le passé, la coopération est indispensable.»

En 1997 et 1998, de grandes organisations internationales gouvernementales mettent leurs bases terminologiques spécialisées en accès libre sur le web pour les linguistes du monde entier. C’est le cas par exemple pour la base ILOTERM quadrilingue (allemand, anglais, espagnol, français) de l’Organisation internationale du Travail (OIT), la base TERMITE quadrilingue (anglais, espagnol, français, russe) de l’Union internationale des télécommunications (UIT) et la base WHOTERM trilingue (anglais, espagnol, français) de l’Organisation mondiale de la santé (OMS).

Eurodicautom, la base terminologique en ligne de la Commission européenne, est d’abord disponible en interne avant le lancement de son site web gratuit en 1997. Eurodicautom est une base terminologique multilingue de termes économiques, scientifiques, techniques et juridiques permettant de combiner entre elles les onze langues officielles de l’Union européenne (allemand, anglais, danois, espagnol, finnois, français, grec, hollandais, italien, portugais, suédois) et le latin.

Un projet de base terminologique plus vaste est évoqué dès 1999 pour fusionner le contenu de toutes les bases terminologiques de l’Union européenne et renforcer ainsi la coopération inter-institutionnelle. Outre la Commission européenne, les partenaires de ce projet sont le Parlement européen, le Conseil de l’Union européenne, la Cour de justice, la Cour des comptes européenne, le Comité économique et social européen, le Comité des régions, la Banque européenne d’investissement, la Banque centrale européenne et le Centre de traduction des organes de l’Union européenne.

Fort de ses 120.000 consultations par jour, Eurodicautom annonce fin 2003 sa fermeture provisoire et son intégration future dans une base terminologique plus vaste qui ne comporterait plus douze langues mais une vingtaine de langues, du fait de l’élargissement de l’Union européenne prévu l’année suivante, avec 25 pays membres en mai 2004 (et 27 pays membres en janvier 2007).

La nouvelle base IATE (InterActive Terminology for Europe — Terminologie interactive pour l’Europe) voit le jour au printemps 2004, d’abord en interne pour les institutions de l’Union européenne, puis sur le web en juin 2007 avec accès libre et gratuit pour un usage mondial. À cette date, IATE comprend 1,4 million d’entrées dans les 23 langues officielles de l’Union européenne (allemand, anglais, bulgare, danois, espagnol, estonien, finnois, français, grec, hongrois, irlandais, italien, letton, lituanien, maltais, néerlandais, polonais, portugais, roumain, slovaque, slovène, suédois, tchèque) et le latin.

Le site web de IATE est administré par le Centre de traduction des organes de l’Union européenne (situé dans la ville de Luxembourg) pour le compte des partenaires du projet. Selon la brochure en ligne de IATE, elle aussi disponible en 23 langues: «Les termes sont introduits dans la base de données par les terminologues et les traducteurs de l’Union européenne sur la

base des informations fournies par les traducteurs, les administrateurs, les juristes-linguistes, les experts et d’autres sources fiables.» Selon la même brochure, IATE comprend 8,4 millions de termes dans 23 langues en 2010, dont 540.000 abréviations et 130.000 expressions.

La traduction automatique

Le but d’un logiciel de traduction est d’analyser le texte dans la langue source (texte à traduire) et de générer automatiquement le texte correspondant dans la langue cible (texte traduit), en utilisant des règles précises pour le transfert de la structure grammaticale.

L’Association européenne pour la traduction automatique (EAMT) explique en 1998 sur son site web: «Il existe aujourd’hui un certain nombre de systèmes produisant un résultat qui, s’il n’est pas parfait, est de qualité suffisante pour être utile dans certaines applications spécifiques, en général dans le domaine de la documentation technique. De plus, les logiciels de traduction, qui sont essentiellement destinés à aider le traducteur humain à produire des traductions, jouissent d’une popularité croissante auprès des organismes professionnels de traduction.»

À la même date, un historique de la traduction automatique est présent sur le site de Globalink, société spécialisée dans les produits et services de traduction. Voici cet historique résumé en trois paragraphes.

La traduction automatique et le traitement de la langue naturelle font leur apparition à la fin des années 1930, et progressent ensuite de pair avec l’évolution de l’informatique quantitative. Pendant la deuxième guerre mondiale, le développement des premiers ordinateurs programmables bénéficie des progrès de la cryptographie et des efforts faits pour tenter de fissurer les codes secrets allemands et autres codes de guerre. Le secteur émergent des technologies de l’information continue ensuite de s’intéresser de près à la traduction et à l’analyse du texte en langue naturelle. Les recherches portent surtout sur la traduction littérale, à savoir la traduction mot à mot sans prise en compte des règles linguistiques. Un projet de l’Université de Georgetown (États-Unis) débuté en 1950 représente la première tentative systématique visant à créer un système de traduction automatique utilisable de l’anglais vers le russe. Des recherches similaires sont menées en Europe.

En 1965, les progrès rapides en linguistique théorique culminent avec la publication de l’ouvrage «Aspects de la théorie syntaxique» de Noam Chomsky, qui propose de nouvelles définitions pour la phonologie, la morphologie, la syntaxe et la sémantique du langage humain. En 1966, un rapport officiel américain donne une estimation prématurément négative des systèmes de traduction automatique, mettant fin au financement et à l’expérimentation dans ce domaine pour la décennie suivante.

Il faut attendre la fin des années 1970 pour que des expériences sérieuses soient à nouveau entreprises, parallèlement aux progrès de l’informatique et des technologies des langues. Cette période voit le développement de systèmes de transfert d’une langue à l’autre et le lancement des premières tentatives commerciales. Des sociétés comme Systran et Metal sont persuadées de

la viabilité et de l’utilité d’un tel marché. Elles lancent des produits et des services de traduction automatique reliés à un serveur central. Mais les problèmes restent nombreux, par exemple les coûts élevés en recherche et développement, un énorme travail lexicographique, la difficulté de proposer de nouvelles combinaisons de langues et l’inaccessibilité de tels systèmes pour l’utilisateur moyen.

Dans «Web embraces language translation» (Le web accueille la traduction des langues), un article de ZDNN (ZD Network News) paru le 21 juillet 1998, la journaliste Martha L. Stone explique: «Parmi les nouveaux produits d’un secteur de traduction représentant 10 milliards de dollars US, on trouve les logiciels de traduction instantanée de sites web, groupes de discussion, courriers électroniques et intranets d’entreprise. Les principales sociétés de traduction se mobilisent pour saisir les opportunités de ce marché. Voici quelques exemples. Systran s’est associé avec AltaVista pour produire Babel Fish, avec 500.000 à 600.000 visiteurs quotidiens et environ un million de traductions par jour, des traductions qui vont des recettes de cuisine à des pages web complètes. 15.000 sites environ ont un lien vers Babel Fish, qui peut traduire [de l’anglais] vers le français, l’italien, l’allemand, l’espagnol et le portugais, et vice versa. (…)»

«Alis a mis au point le système de traduction du Los Angeles Times qui doit bientôt être lancé sur le site et qui proposera des traductions [de l’anglais] vers l’espagnol et le français, et plus tard le japonais. D’un clic de souris, une page web complète peut être traduite dans la langue désirée. Globalink propose des logiciels, des systèmes de traduction de pages web, un service de messagerie électronique gratuit et des logiciels permettant de traduire les textes des groupes de discussion. Cependant, alors que ces systèmes de traduction automatique deviennent populaires dans le monde entier, les directeurs des sociétés qui les développent admettent qu’ils ne peuvent répondre à toutes les situations. Les porte-paroles de Globalink, Alis et Systran utilisent des expressions comme “pas parfait” et “approximatif” lorsqu’ils décrivent la qualité des traductions, et précisent bien que les phrases soumises à la traduction doivent être simples, grammaticalement correctes et sans tournures idiomatiques.»

En mars 2001, IBM lance un produit professionnel haut de gamme, le WebSphere Translation Server (serveur de traduction). Ce serveur traduit instantanément en plusieurs langues (allemand, anglais, chinois, coréen, espagnol, français, italien, japonais) les pages web, courriels et chats (dialogues en direct). Il interprète 500 mots à la seconde et permet l’ajout de vocabulaires spécifiques. D’autres langues viennent ensuite s’ajouter aux huit langues de base. Et les sociétés Systran, Alis Technologies, Lernout & Hauspie (qui a racheté Globalink) et Softissimo développent des logiciels, produits et services ciblant trois types de clients: les sociétés localisant leurs sites web, les professionnels des langues et le grand public.

La traduction assistée par ordinateur

Un logiciel de traduction automatique (TA) analyse le texte dans la langue à traduire (langue source) et génère automatiquement le texte dans la langue désirée (langue cible), en utilisant des règles précises pour le transfert de la structure grammaticale. L’être humain n’intervient pas

dans ce processus, contrairement à la traduction assistée par ordinateur (TAO), qui implique une interaction entre l'être humain et la machine.

Des sociétés lancent des logiciels de traduction assistée par ordinateur à destination des traducteurs professionnels. Le plus populaire est Wordfast, lancé dès 1999 par Yves Champollion, avec gestion de la terminologie en temps réel et contrôle typographique. Wordfast est compatible avec d'autres logiciels tels que le WebSphere Translation Server d'IBM et SDL Trados. Une version simplifiée de Wordfast est téléchargeable gratuitement, avec un manuel d'utilisation disponible en 16 langues et une application disponible sur le web.

En 2010, Wordfast est le numéro un mondial des logiciels de traduction utilisables aussi bien sur plateforme Windows (Microsoft) que sur plateforme Mac OS (Apple), et le numéro deux mondial en nombre de ventes (après SDL Trados), avec 20.000 clients dans le monde, dont les Nations Unies, la NASA (National Aeronautics and Space Administration — Administration nationale de l'aéronautique et de l'espace), Sony, Coca-Cola et l'éditeur McGraw-Hill.

Selon Tim McKenna, professeur de mathématiques et écrivain, interviewé en octobre 2000: «Lorsque la qualité des logiciels sera suffisante pour que les gens puissent discuter sur le web en temps réel dans différentes langues, nous verrons tout un monde s'ouvrir à nous. Les scientifiques, les hommes politiques, les hommes d'affaires et bien d'autres groupes seront à même de communiquer immédiatement entre eux sans l'intermédiaire de médiateurs ou traducteurs.»

L'étape suivante pourrait être celle de la «transparence transculturelle et transnationale» décrite en septembre 1998 par Randy Hobler, consultant en marketing internet des produits et services de traduction : «Nous arriverons rapidement au point où une traduction très fidèle du texte et de la parole sera si commune qu'elle pourra faire partie des plateformes ou même des puces. À ce stade, lorsque le développement de l'internet aura atteint sa vitesse de croisière, lorsque la fidélité de la traduction atteindra plus de 98% et lorsque les différentes combinaisons de langues possibles auront couvert la grande majorité du marché, la transparence de la langue (toute communication d'une langue à une autre) sera une vision trop restrictive pour ceux qui vendent cette technologie. Le développement suivant sera la “transparence transculturelle et transnationale” dans laquelle les autres aspects de la communication humaine, du commerce et des transactions au-delà du seul langage entreront en scène. Par exemple, les gestes ont un sens, les mouvements faciaux ont un sens, et ceci varie en fonction des sociétés. La lettre O faite avec le pouce et l'index signifie “OK” aux États-Unis alors qu'en Argentine c'est un geste obscène.»

«Lorsque se produira l'inévitable développement de la vidéoconférence multilingue multimédia, il sera nécessaire de corriger visuellement les gestes. Le Media Lab du MIT, Microsoft et bien d'autres travaillent à la reconnaissance informatique des expressions faciales, l'identification des caractéristiques biométriques du visage, etc. Il ne servira à rien à un homme d'affaires américain de faire une excellente présentation à un Argentin lors d'une vidéoconférence multilingue sur le web, avec son discours traduit dans un espagnol argentin parfait, s'il fait en même temps le geste O avec le pouce et l'index. Les ordinateurs pourront intercepter ces types de messages et les corriger visuellement.»

«Les cultures diffèrent de milliers de façons, et la plupart d'entre elles peuvent être modifiées par voie informatique lorsqu'on passe de l'une à l'autre. Ceci inclut les lois, les coutumes, les habitudes de travail, l'éthique, le change monétaire, les différences de taille dans les vêtements, les différences entre le système métrique et le système de mesure anglophone, etc. Les entreprises dynamiques répertorieront et programmeront ces différences, et elles vendront des produits et services afin d'aider les habitants de la planète à mieux communiquer entre eux. Une fois que ceux-ci seront largement répandus, ils contribueront réellement à une meilleure compréhension à l'échelle internationale.»

La traduction automatique gratuite en ligne

Le moteur de recherche AltaVista lance fin 1997 un service de traduction gratuit de l'anglais vers cinq autres langues (allemand, espagnol, français, italien, portugais) et vice versa, dénommé Babel Fish ou AltaVista Translation. La page web originale et la traduction apparaissent en vis-à-vis à l'écran. On peut également traduire n'importe quel texte court en faisant un copier-coller dans l'interface web. Bien qu'ayant ses limites avec un texte traduit très approximatif, ce service est immédiatement plébiscité par les 12 millions d'usagers que compte le web de l'époque, dont un nombre croissant d'usagers non anglophones, et contribue grandement au plurilinguisme du web.

Alimenté par des dictionnaires multilingues comprenant 2,5 millions de termes, Babel Fish est l'oeuvre de Systran, société spécialisée dans le traitement automatique des langues. Selon le site web de Systran: «Un logiciel de traduction automatique traduit une langue naturelle dans une autre langue naturelle. La traduction automatique prend en compte la structure grammaticale de chaque langue et elle utilise des règles pour transférer la structure grammaticale de la langue source (texte à traduire) vers la langue cible (texte traduit). La traduction automatique ne remplace pas et n'est pas destinée à remplacer le traducteur humain.» Beaucoup plus tard, en mai 2008, Babel Fish déménage sur Yahoo! avant d'être remplacé par Bing (Microsoft) en 2012.

Lancé en octobre 2007, Google Traduction (Google Translate) est un service en ligne gratuit qui traduit instantanément un texte ou une page web dans une autre langue, avec un résultat immédiat à l'écran vis-à-vis de l'original. Innovation par rapport à la concurrence, Google Traduction se base sur une analyse statistique pour produire la traduction automatique et non sur une analyse traditionnelle basée sur des règles linguistiques. Google explique aussi que ce nouveau service peut aider l'usager à comprendre le sens général d'un texte en langue étrangère, mais ne propose pas de traductions exactes.

Avant le lancement de Google Traduction, Google utilisait le service de traduction de Systran (du même type que Babel Fish), avec plusieurs étapes pour les langues disponibles, détaillées par Wikipédia: (1) le français, l'allemand et l'espagnol en plus de l'anglais; (2) le portugais et le flamand; (3) l'italien; (4) le chinois simplifié, le japonais et le coréen; (5) l'arabe (avril 2006); (6) le russe (décembre 2006); (7) le chinois traditionnel (février 2007).

Lorsque Google remplace le logiciel de traduction de Systran par son propre logiciel, les étapes sont les suivantes entre 2007 et 2010, toujours selon Wikipédia: (1) toutes les langues disponibles jusqu'ici, avec toutes les combinaisons possibles (octobre 2007); (2) l'hindou (décembre 2007); (3) le bulgare, le croate, le danois, le finlandais, le grec, le néerlandais, le norvégien, le polonais, le roumain, le suédois et le tchèque (mai 2008); (4) le catalan, l'hébreu, l'indonésien, le letton, le lituanien, le philippin, le serbe, le slovaque, le slovène, l'ukrainien et le vietnamien (septembre 2008); (5) l'albanais, l'estonien, le galicien, le hongrois, le maltais, le thaï et le turc (janvier 2009); (6) le perse (juin 2009); (7) l'afrikaans, le biélorusse, le gallois, l'irlandais, l'islandais, le macédonien, le malais, le swahili et le yiddish (août 2009); (8) le créole haïtien (janvier 2010); (9) l'arménien, l'azéri, le basque, le géorgien et l'ourdou (mai 2010); (10) le latin (octobre 2010).

Google Traduction lance en 2009 une lecture par synthèse vocale pour quelques langues, avec l'ajout de langues supplémentaires au fil des mois, ainsi que le choix entre plusieurs traductions pour un même mot en janvier 2011. Fort utile pour les traducteurs (humains), le Google Translator Toolkit (Boîte à outils pour les traducteurs), lancé lui aussi en 2009, leur permet de réviser les traductions générées automatiquement par Google Traduction de l'anglais vers 47 langues cibles, tout en partageant leurs traductions, glossaires et mémoires de traduction. Le service est étendu à d'autres paires de langues ensuite.

Le catalogue des langues vivantes

«Ethnologue: Languages of the World» (Ethnologue: les langues du monde) est un catalogue encyclopédique de toutes les langues vivantes qui crée sa version web gratuite en 1996, avec présentation des langues selon divers critères (nom de la langue, famille linguistique, pays dans lequel la langue est parlée, identifiant de trois lettres, etc.) ainsi que des index et des cartes géographiques. La version en ligne devient payante en 2015 pour financer le travail en amont.

Publié par SIL International depuis 1951, l'Ethnologue est d'abord un catalogue de langues minoritaires avant de s'élargir à toutes les langues vivantes en 1971. Basée à Dallas (États-Unis), l'équipe de chercheurs de SIL International rassemble et organise la masse d'informations glanées et vérifiées une à une sur le terrain par des milliers de linguistes regroupés dans des équipes nationales et/ou linguistiques présentes sur tous les continents.

Barbara Grimes, directrice de publication de l'Ethnologue entre 1971 et 2000, relate en janvier 2000 dans un entretien par courriel: «L'Ethnologue est un catalogue des langues dans le monde, avec des informations sur les pays où ces langues sont parlées, une estimation du nombre de personnes qui les parlent, la famille linguistique à laquelle elles appartiennent, les autres termes utilisés pour ces langues, les noms de dialectes, diverses informations socio-linguistiques et démographiques, les dates des Bibles publiées, un index des noms de langues [Ethnologue Name Index], un index des familles linguistiques [Ethnologue Language Family Index] et enfin des cartes géographiques pour les langues.»

Mais qu’est-ce exactement qu’une langue? Selon le site de l’Ethnologue: «La manière dont chacun choisit de définir une langue dépend des motifs qu’on a d’identifier cette langue comme étant distincte d’une autre. Certains basent la définition d’une langue sur des raisons purement linguistiques. D’autres reconnaissent la nécessité de prendre également en compte des facteurs sociaux, culturels ou politiques. En outre, les locuteurs d’une langue ont souvent leurs propres critères sur l’appropriation d’une langue comme étant la leur. Ces critères sont souvent bien davantage liés à des questions de patrimoine et d’identité qu’aux traits linguistiques de la langue ou des langues en question.»

La 17e édition (2013) de l’Ethnologue recense 7.105 langues vivantes, et publie pour la première fois son édition en ligne gratuite avant son édition imprimée payante. L’Ethnologue annonce aussi une actualisation chaque année et non plus tous les quatre ans, pour être en phase avec le monde rapide dans lequel nous vivons.

Cette 17e édition nous apprend que la langue anglaise reste prédominante dans le monde puisqu’elle est répertoriée dans 125 pays, soit comme langue officielle soit comme langue parlée par un groupe significatif d’immigrants. Sur les 136 familles linguistiques existantes, la plus importante est l’indo-européen, qui regroupe 3 milliards de locuteurs. La Papouasie-Nouvelle-Guinée est le pays qui recense le plus de langues vivantes, avec 836 langues. Suivent l’Indonésie avec 706 langues et le Nigéria avec 522 langues.

Une caractéristique de la base de données de l’Ethnologue depuis 1971 est un identifiant de trois lettres pour chaque langue (par exemple «fra» pour le français), avec inclusion des identifiants dans la version imprimée à partir de la 10e édition (1984).

En 2002, à l’invitation de l’Organisation internationale de normalisation (ISO), SIL International met en chantier une nouvelle norme ISO destinée à remplacer la norme ISO 639-2, publiée en 1998 pour identifier 400 langues (et succéder à la norme ISO 639-1 de 1988). L’identification de 400 langues n’est plus suffisante avec la mondialisation du réseau. SIL International harmonise donc les identifiants utilisés dans l’Ethnologue avec ceux de la norme ISO 639-2, en intégrant aussi les identifiants des langues mortes et artificielles utilisés par la Linguist List, grande liste de diffusion à destination des linguistes.

Publiée en 2007, la norme ISO 639-3 attribue un identifiant de trois lettres à près de 7.500 langues vivantes, mortes et artificielles, y compris les langues des signes. SIL International est également désigné comme l’organisme responsable de la gestion du cycle annuel des modifications et des mises à jour.

Les langues minoritaires

Guy Antoine, un informaticien né à Haïti et installé à New York, crée le site Windows on Haiti en avril 1998 afin de promouvoir le créole haïtien, parlé non seulement à Haïti mais aussi en République dominicaine, aux États-Unis et au Canada.

Il raconte en juin 2001 dans un entretien par courriel: «Que sont les Haïtiens sans le kreyôl (créole pour les non-initiés), une langue qui s'est développée et qui a permis de souder entre elles diverses tribus africaines transplantées à Haïti pendant la période de l'esclavage? Cette langue représente de manière la plus palpable l'unité de notre peuple. Cependant elle est surtout une langue parlée et non écrite. À mon avis, le web va changer cet état de fait plus qu'aucun autre moyen traditionnel de diffusion d'une langue. Mon site souhaite d'une part être une source d'information majeure sur la culture haïtienne, d'autre part contrer les images continuellement négatives que les médias traditionnels donnent d'Haïti. Je voulais aussi montrer la diversité de la culture haïtienne dans des domaines tels que l'art, l'histoire, la cuisine, la musique, la littérature et la vie traditionnelle.»

Quelles sont les langues du site? «La langue principale est l'anglais, mais le kreyôl est bien présent dans les forums de discussion. On trouve aussi des documents sur Haïti en français et dans l'ancien créole colonial, et je suis prêt à publier d'autres documents en espagnol et dans diverses langues. Je ne propose pas de traductions, mais le multilinguisme est effectif sur ce site, et je pense qu'il deviendra de plus en plus la norme sur le web. Le premier forum regroupe des discussions générales sur toutes sortes de sujets, mais en fait ces discussions concernent principalement les problèmes socio-politiques qui agitent Haïti. Le deuxième forum est uniquement réservé aux débats sur les normes d'écriture du kreyôl. Ces débats sont assez animés, et un certain nombre d'experts linguistiques y participent. Le caractère exceptionnel de ces forums est qu'ils ne sont pas académiques.»

Le site Indigenous Tweets (Tweets autochtones) est créé en mars 2011 par Kevin Scannell, informaticien et professeur à l'Université Saint-Louis (États-Unis). Ce projet mené sur son temps libre consiste à recenser les tweets dans les langues minoritaires et autochtones, dans le but de renforcer la présence en ligne des communautés parlant ses langues. Pour trouver ces langues sur Twitter, Kevin Scannell conçoit et paramètre An Crúbadán, un logiciel statistique crawlant le web. Indigenous Tweets débute avec 35 langues en mars 2011 et compte 71 langues trois semaines plus tard, 144 langues en mars 2013 et 184 langues en octobre 2017.

La page d'accueil est une liste des langues minoritaires actives sur Twitter. L'usager cherche la langue qui l'intéresse et clique sur la ligne correspondante, ce qui l'amène vers une nouvelle page recensant les usagers de Twitter dans cette langue (500 usagers au maximum) et les statistiques correspondantes pour chaque usager, entre autres le nombre de tweets, le nombre d'abonnés, le pourcentage de tweets dans la langue donnée (certains tweetent à la fois dans une langue minoritaire et dans une langue globale) et la date du tweet le plus récent. De nombreux projets ont débuté par un premier contact entre deux personnes sur Twitter.

Kevin Scannell explique en mars 2011 sur son blog que le but premier d'Indigenous Tweets est d'aider à créer des communautés de langues par le biais de Twitter, et de permettre aux gens parlant une langue minoritaire ou autochtone de se rencontrer plus facilement dans la vaste mer de l'anglais, du français, de l'espagnol ou d'autres langues globales dominant Twitter. Les principales langues minoritaires sur Twitter sont le créole haïtien, le basque et l'irlandais. Mais, d'après Kevin Scannell, même les locuteurs des langues basque et irlandaise, qui ont des communautés en ligne très actives, ont été surpris de voir le nombre de gens qui tweetaient dans ces langues.

Kevin Scannell créé un deuxième site, Indigenous Blogs (Blogs autochtones), en septembre 2011, cette fois pour recenser les blogs rédigés dans des langues minoritaires et autochtones. Il recense d'abord les blogs hébergés par Blogspot (qui héberge aussi son propre blog), WordPress et Tumblr. La présentation d'Indigenous Blogs est semblable à celle d'Indigenous Tweets, à savoir une liste de toutes les langues sur la page d'accueil puis une page par langue. Indigenous Blogs recense des blogs dans 50 langues en septembre 2011, 74 langues en mars 2013 et 85 langues en octobre 2017.

Les langues menacées

L'UNESCO (Organisation des Nations Unies pour l'éducation, la science et la culture) lance en 2010 son «Atlas interactif des langues en danger dans le monde» en trois langues (anglais, français, espagnol). La version en ligne gratuite est complémentaire de la 3e édition imprimée (2010) sous la direction de Christopher Moseley. Les deux premières éditions imprimées dataient respectivement de 1996 et 2001 et n'avaient pas de version en ligne.

L'atlas comprend 2.473 langues en juin 2011 et 2.464 langues en juillet 2017, avec recherche possible selon divers critères (pays ou région, nom de langue, nombre de locuteurs, niveau de vitalité d'une langue, code ISO 639-3). Les noms des langues sont transcrits en français, en anglais et en espagnol sans oublier leurs noms alternatifs (variantes orthographiques, dialectes ou noms en caractères non latins).

Le «Rapport de l'UNESCO sur la vitalité et le danger de disparition des langues» établit six niveaux de vitalité pour une langue: sûre, vulnérable, en danger, sérieusement en danger, en situation critique, éteinte. (1) «Sûre» signifie que la langue est parlée par toutes les générations et que la transmission intergénérationnelle est ininterrompue. Les langues concernées ne sont donc pas incluses dans l'atlas. (2) «Vulnérable» signifie que la plupart des enfants parlent la langue, mais qu'elle est restreinte à certains lieux, par exemple le domicile. (3) «En danger» signifie que les enfants n'apprennent plus la langue comme langue maternelle chez eux. (4) «Sérieusement en danger» signifie que la langue est parlée par les grands-parents. Si la génération des parents peut la comprendre, les parents ne la parlent pas entre eux ou avec leurs enfants. (5) «En situation critique» signifie que les locuteurs les plus jeunes sont les grands-parents et leurs ascendants, et qu'ils ne parlent la langue que partiellement et peu fréquemment. (6) «Éteinte» signifie qu'il n'y a plus de locuteurs. L'atlas inclut les langues éteintes depuis les années 1950.

À quel moment une langue est-elle considérée comme en péril? Comme expliqué sur le site de l'atlas: «Une langue est en péril lorsque ses locuteurs cessent de l'utiliser, réservent son usage à des domaines de plus en plus restreints, emploient un moins grand nombre de registres ou de styles de parole, et/ou arrêtent de la transmettre à la génération suivante. Aucun facteur ne détermine à lui seul si une langue est en danger.»

Selon les experts de l'UNESCO, il importe de considérer les neuf critères suivants: (1) la transmission de la langue d'une génération à l'autre, (2) le nombre absolu de locuteurs, (3) le taux de locuteurs par rapport à l'ensemble de la population, (4) l'utilisation de la langue dans les différents lieux publics et privés, (5) la réactivité d'une langue face aux nouveaux sujets et médias, (6) l'existence de matériel d'apprentissage et d'enseignement de la langue, (7) les attitudes et politiques linguistiques du gouvernement et des institutions, y compris l'usage officiel et le statut officiel de la langue, (8) les attitudes des membres de la communauté concernée vis-à-vis de leur propre langue, (9) le type et la qualité de la documentation disponible dans cette langue.

Quels sont les facteurs de disparition d'une langue? Selon les mêmes experts: «Une langue disparaît lorsqu'elle n'a plus de locuteurs ou que ceux-ci se mettent à parler une autre langue — en général une langue de plus grande importance utilisée par un groupe plus puissant. Les langues sont menacées par des forces externes telles qu'une domination militaire, économique, religieuse, culturelle ou éducative, ou par des forces internes comme l'attitude négative d'une population à l'égard de sa propre langue. Aujourd'hui, les migrations accrues et l'urbanisation rapide s'accompagnent souvent de la perte des modes de vie traditionnels et d'une forte pression en faveur de l'utilisation d'une langue dominante qui est nécessaire — ou perçue comme telle — à une vraie participation à la vie civique et au progrès économique.»

L'atlas de l'UNESCO considère par exemple le gaélique écossais comme une langue «sérieusement en danger». D'après le recensement de 2011, moins de 60.000 personnes parlent le gaélique, soit un peu plus de 1% de la population de l'Écosse, et 92.000 personnes comprennent la langue. Ces chiffres sont très inférieurs à ceux du recensement de 1901, qui comptaient 200.000 personnes parlant le gaélique, soit 4,5% de la population.

La situation était très différente par le passé. Pendant de nombreux siècles, tout le monde parle le gaélique en Écosse et en Irlande et les universitaires diffusent leurs écrits en gaélique dans toute l'Europe. Au fil des siècles, l'anglais devient peu à peu la langue dominante, y compris sur les îles occidentales écossaises, malgré la présence du gaélique écossais en tant que première langue communautaire. La culture gaélique renaît au début du 19e siècle, sous forme de poésie, de prose et de musique. Entre les deux guerres mondiales, la radio diffuse les actualités en gaélique et on apprend la langue à l'école. De nos jours, nombre de romans sont publiés en gaélique. Radio nan Gàidheal émet en gaélique depuis les années 1980 et la chaîne de télévision ALBA diffuse ses émissions en gaélique depuis le début des années 2000. Toutes deux sont présentes sur le web, ce qui a boosté leur audience.

Qu'en est-il du gaélique sur le web? Opera est le premier navigateur web offrant une interface en gaélique, en 2001, suivi de Firefox (Mozilla) et de Google Chrome, grâce au patient travail de Michael Bauer, traducteur indépendant, sur son temps libre, avec l'aide de sa collègue GunChleoc (pseudonyme qui signifie «une femme» en gaélique écossais). Michael Bauer et GunChleoc sont aussi les auteurs des versions gaéliques de Thunderbird (messagerie de Mozilla), Lightning (calendrier de Mozilla), OpenOffice, LibreOffice, auxquels s'ajoutent les versions gaéliques du VLC media player, du jeu Freeciv (version open source du jeu Civilisation) et de Accentuate.us (logiciel permettant d'insérer automatiquement les accents). Michael Bauer est également l'auteur (avec Kevin Scannell) du correcteur d'orthographe An Dearbhair Beag.

Wikipédia a sa version gaélique, Uicipeid. Le gaélique écossais dispose de trois dictionnaires en ligne. Le premier dictionnaire est Stòr-dàta, une liste de mots gérée par l'Institut Sabhal Mòr Ostaig (situé sur l'île de Skye, en Écosse), dont tous les cours sont dispensés en gaélique écossais. Le deuxième dictionnaire est le Dwelly, le fameux dictionnaire gaélique datant de 1911, qui est au gaélique ce que l'Oxford English Dictionary est à l'anglais. Cette version numérisée a demandé dix ans de travail à Michel Bauer, avec l'aide de son collègue Will Robertson. Le troisième dictionnaire est Am Faclair Beag, qui signifie «petit dictionnaire» mais qui est en fait un vaste dictionnaire regroupant le Dwelly de 1911 et des données plus modernes, toujours grâce au patient travail de Michael Bauer et Will Robertson.

Et les projets ne manquent pas. D'après Michael Bauer, interviewé en octobre 2015, une archive en ligne serait très utile pour les projets de localisation, avec une mémoire de traduction commune, ce qui éviterait de devoir retraduire indéfiniment les mêmes termes et segments de phrase. Si on avait la possibilité de traduire les logiciels à partir d'une archive en ligne commune, par exemple un genre de méta-Pootle (serveur pour les projets communs de localisation), tout le monde en serait bénéficiaire, non seulement pour le gaélique mais pour l'ensemble des langues minoritaires.

De nombreuses langues minoritaires, autochtones et menacées ont encore besoin de dictionnaires, de grammaires et de glossaires. Certaines langues ont même besoin de technologies de base telles que configuration de clavier et correcteur d'orthographe. La tâche est à la fois passionnante et titanesque.

Nombre de locuteurs de ces langues se battent pour garder leur langue et leur culture en vie. De plus en plus de groupes linguistiques se tournent vers le web, devenu un outil majeur pour la revitalisation d'une langue. Des milliers de gens bloguent dans leur langue natale et utilisent les réseaux sociaux tels que Facebook et Twitter dans ces langues.

Les réseaux sociaux ont permis à des communautés parfois éparpillées de nouer des liens et d'utiliser leur langue de manière naturelle. Ils ont permis aussi de toucher les jeunes, un groupe démographique essentiel pour la revitalisation d'une langue, puisqu'une langue menacée est presque toujours une langue parlée seulement par les anciens. Plus les jeunes parlent et écrivent une langue menacée, moins elle risque de disparaître.

Chronologie

[année-mois]
1963: Le premier système d'encodage informatique est l'ASCII (American Standard Code for Information Interchange).
1971-07: L'eText #1 de Michael Hart touche les cent usagers du pré-internet. Le Projet Gutenberg est né.
1974: Vinton Cerf et Robert Kahn créent les protocoles de communication de l'internet.
1976: Le Congrès américain durcit sa législation avec un copyright de 50 ans après le décès de

l’auteur.
1977: L’IFLA (International Federation of Library Associations) crée l’UNIMARC en tant que format bibliographique commun pour le catalogage en bibliothèque.
1983: Après avoir été un réseau gouvernemental et académique américain, l’internet prend son envol à l’international.
1984: À l’instigation de Richard Stallman, le copyleft est institué pour les logiciels afin de permettre leur libre réutilisation et il est formalisé avec la GPL (General Public License).
1990: Tim Berners-Lee invente le World Wide Web et offre son invention au monde.
1991-01: L’Unicode Consortium est fondé pour développer l’Unicode, un nouveau système d’encodage permettant d’afficher toutes les langues de la planète à l’écran.
1992: Le Projekt Runeberg est la première collection numérique suédoise offrant des livres du domaine public.
1992: Vinton Cerf fonde l’Internet Society (ISOC) pour coordonner le développement de l’internet.
1992: Paul Southworth crée les Etext Archives pour accueillir des textes électroniques de tous ordres.
1993: John Mark Ockerbloom crée The Online Books Page pour offrir un point commun aux livres en accès libre sur le web.
1993-04: L’Association des bibliophiles universels (ABU) crée la première collection numérique française de livres du domaine public.
1993-06: Adobe lance le format PDF (Portable Document Format) et l’Acrobat Reader.
1993-07: John Labovitz crée l’E-Zine-List pour recenser les zines électroniques.
1993-11: Le web voit arriver Mosaic, premier navigateur grand public.
1994: Le Netscape Navigator succède à Mosaic comme navigateur du web.
1994: Le Projekt Gutenberg-DE est la première collection numérique allemande de livres du domaine public.
1994: Michel Martin crée Travlang, un répertoire de dictionnaires de langues en ligne pour les voyageurs.
1994: Pierre Perroud crée Athena, une collection numérique suisse de livres du domaine public.
1994-02: La bibliothèque publique d’Helsinki (Finlande) est la première à créer un site web.
1994-05: Tyler Chambers crée la Human-Languages Page (H-LP) pour recenser les ressources linguistiques en ligne.
1994-07: Les usagers de l’internet dont la langue maternelle n’est pas l’anglais atteignent les 5%.
1994-10: Le World Wide Consortium (W3C) est fondé pour développer les protocoles communs du web.
1995: La grande presse imprimée débute sa présence sur le web.
1995: Le WorldWide Language Institute (WWLI) crée NetGlos, un glossaire coopératif multilingue des termes liés à l’internet et à l’informatique.
1995: Microsoft lance son propre navigateur, l’Internet Explorer.
1995: Robert Beard crée A Web of Online Dictionaries (WOD) pour répertorier les dictionnaires gratuits en ligne.
1995: Tyler Chambers lance l’Internet Dictionary Project (IDP) pour créer des dictionnaires coopératifs bilingues en accès libre.
1995-07: Jeff Bezos lance la librairie en ligne Amazon.com, pionnier du cybercommerce.
1996: L’Ethnologue, catalogue encyclopédique des langues vivantes, débute sa version web

gratuite.
1996-04: L’Internet Archive est fondée par Brewster Kahle pour archiver le web à l’intention des générations présentes et futures.
1996-04: Robert Ware lance OneLook Dictionaries, un moteur de recherche commun à de nombreux dictionnaires en ligne.
1996-12: Le Traité de l’OMPI (Organisation mondiale de la propriété intellectuelle) sur le droit d’auteur est signé par les pays membres pour réguler les éditions numériques.
1997: Le premier blog voit le jour sur le web.
1997: Hachette met en ligne son Dictionnaire universel francophone, avec accès libre et gratuit.
1997-01: Le Bureau international du travail (BIT) organise son premier colloque sur la convergence multimédia.
1997-01: Les bibliothèques nationales européennes créent Gabriel, un site web offrant un point d’accès commun à leurs services en ligne.
1997-04: Le web mondial compte un million de sites.
1997-05: La British Library met son catalogue en ligne avec accès libre et gratuit.
1997-08: O’Reilly publie le livre “Pour un web multilingue” de Yoshi Mikami en japonais, et le traduit en anglais, en allemand et en français l’année suivante.
1997-09: L’Internet Bookshop (Royaume-Uni) débute la vente de livres provenant des États-Unis.
1997-10: La Bibliothèque nationale de France (BnF) lance sa bibliothèque numérique Gallica avec 3.000 ouvrages du 19e siècle.
1997-12: La société italienne de traduction Logos met tous les outils linguistiques destinés à ses traducteurs (dictionnaires, glossaires, grammaires, conjugaisons) en accès libre.
1997-12: Le moteur de recherche AltaVista lance Babel Fish, son service de traduction automatique gratuit.
1997-12: L’internet mondial compte 70 millions d’usagers, soit 1,7% de la population mondiale.
1998: Le projet franco-américain ARTFL met en ligne la base de données correspondant au premier volume de l’Encyclopédie de Diderot.
1998-07: Les usagers de l’internet dont la langue maternelle n’est pas l’anglais atteignent les 20%.
1998-10: Le Digital Millennium Copyright Act (DMCA) est entériné par le Congrès américain et fait passer le copyright de 50 ans à 70 ans après la mort de l’auteur.
1999: Michael Kellogg crée WordReference.com pour proposer gratuitement des dictionnaires en ligne bilingues et des forums de discussion pour les linguistes.
1999-09: Le format Open eBook (OeB) est créé en tant que format standard du livre numérique (à côté du PDF).
1999-12: L’Encyclopaedia Britannica met en ligne sa version numérique Britannica.com, d’abord gratuite puis payante.
1999-12: L’Encyclopaedia Universalis propose aussi une version en ligne d’abord gratuite puis payante.
1999-12: WebEncycIo (éditions Atlas) est la première grande encyclopédie francophone gratuite conçue directement pour le web.
2000-01: Le wiki — site participatif — devient populaire et sert de base à Wikipédia l’année suivante.
2000-01: Le Million Book Project veut proposer un million de livres gratuits dans plusieurs langues sur le web.

2000-02: Robert Beard cofonde yourDictionary.com, grand portail de dictionnaires et autres outils linguistiques en ligne.
2000-03: L'Oxford University Press lance la version en ligne de l'Oxford English Dictionary (OED), avec consultation payante.
2000-03: L'internet mondial compte 300 millions d'usagers, soit 5% de la population mondiale.
2000-07: Les usagers de l'internet dont la langue maternelle n'est pas l'anglais atteignent les 50%.
2000-09: Le Grand dictionnaire terminologique (GDT) du Québec est un dictionnaire en ligne bilingue français-anglais en accès libre et gratuit.
2000-10: Charles Franks crée Distributed Proofreaders pour corriger à plusieurs les versions numérisées des livres du domaine public avant de les intégrer au Projet Gutenberg.
2000-10: La Public Library of Science (PLOS) est fondée pour promouvoir la diffusion libre des revues scientifiques et médicales.
2000-11: La version numérisée de la Bible de Gutenberg originale est disponible sur le site web de la British Library.
2001: Lawrence Lessig conçoit la Creative Commons pour favoriser la diffusion d'oeuvres de tous ordres sur l'internet tout en protégeant le droit d'auteur.
2001-01: Jimmy Wales et Larry Sanger créent Wikipédia, grande encyclopédie collaborative en ligne gratuite.
2001-03: IBM lance le WebSphere Translation Server, un serveur de traduction automatique pouvant traiter huit langues.
2001-05: La Commission européenne entérine la directive EUCD (Directive du droit d'auteur dans l'Union européenne) en vue d'harmoniser les législations nationales sur le droit d'auteur.
2001-10: Avec 30 milliards de pages archivées depuis 1996, l'Internet Archive lance la Wayback Machine pour consulter un site web à différentes dates.
2002- 02: L'Initiative de Budapest pour un accès ouvert (BOAI – Budapest Open Access Initiative) est signée pour promouvoir l'accès ouvert (libre) à la littérature scientifique.
2002-03: L'Oxford University Press lance l'Oxford Reference Online (ORO), une encyclopédie conçue directement pour le web, avec abonnement payant.
2002-12: Les premières licences-type Creative Commons sont publiées par l'organisme du même nom.
2003-09: Le Massachusetts Institute of Technology (MIT) crée son OpenCourseWare pour offrir le matériel pédagogique de ses cours en accès libre et gratuit.
2003-10: La Public Library of Science (PLOS) devient un éditeur de revues scientifiques et médicales gratuites en ligne, dont tous les articles sont sous licence Creative Commons.
2003- 12: Une licence Creative Commons est utilisée par un million d'oeuvres sur l'internet.
2004: L'éditeur Tim O'Reilly lance le terme web 2.0, repris ensuite dans le monde entier pour définir un web participatif.
2004- 01: La Bibliothèque européenne (European Library) prend la suite de Gabriel en tant que portail des bibliothèques nationales européennes.
2004-02: Mark Zuckerberg crée Facebook pour un public étudiant avant de conquérir le monde.
2004-05: Le nombre de langues officielles de l'Union européenne passe de onze langues (plus le latin) à vingt langues suite à son élargissement.
2004-10: Google lance Google Print avant de le rebaptiser Google Books (Google Livres).
2005-04: L'IDPF (International Digital Publishing Forum) succède à l'OeBF (Open eBook

Forum) pour gérer le format OeB (Open eBook).
2005-10: L'Internet Archive fonde l'Open Content Alliance (OCA) pour fédérer plusieurs partenaires et proposer une bibliothèque numérique publique mondiale.
2005-12: Le Massachusetts Institute of Technology (MIT) crée l'OpenCourseWare Consortium pour diffuser en ligne gratuitement le matériel pédagogique des cours d'autres universités.
2005-12: L'internet mondial compte un milliard d'usagers, soit 15,7% de la population mondiale.
2006: La Public Library of Science (PLOS) lance PLOS ONE, revue académique en ligne ouverte à tout sujet scientifique ou médical.
2006-03: La Commission européenne débute son projet de bibliothèque numérique européenne, qui deviendra Europeana deux ans plus tard.
2006-06: Twitter est un outil de micro-blogging pour diffuser des messages de 140 caractères.
2006-08: Google lance Google Books (Google Livres) en remplacement de Google Print.
2006-08: La loi DADVSI (Droit d'auteur et droits voisins dans la société de l'information) est promulguée en France.
2006-08: WorldCat, catalogue collectif mondial géré par OCLC, lance sa version gratuite sur le site web Worldcat.org.
2006-11: Le web mondial compte cent millions de sites.
2006-12: Gallica, bibliothèque numérique de la Bibliothèque nationale de France (BnF), offre 90.000 ouvrages et 80.000 images libres de droits.
2007-01: Le nombre de langues officielles de l'Union européenne passe de 20 langues à 23 langues, avec l'ajout du bulgare, de l'irlandais et du roumain.
2007-02: Creative Commons publie les versions 3.0 de ses licences, avec licence internationale et compatibilité avec des licences similaires (copyleft, GPL et autres).
2007-03: Larry Sanger lance Citizendium en tant qu'encyclopédie en ligne collaborative gratuite gérée par des experts.
2007-04: Le portail yourDictionary.com répertorie 2.500 dictionnaires et grammaires dans 300 langues.
2007-05: L'Encyclopedia of Life voit le jour en tant qu'encyclopédie collaborative en ligne visant à répertorier toutes les espèces animales et végétales connues.
2007-06: La Commission européenne lance la version publique de sa base terminologique multilingue IATE (InterActive Terminology for Europe).
2007-09: L'International Digital Publishing Forum (IDPF) publie le format EPUB, qui remplace le format OeB.
2007-10: Google lance son propre service de traduction automatique en ligne sous le nom de Google Translate (Google Traduction), après avoir utilisé un service de Systran.
2007-12: L'Unicode (créé en 1991) supplante définitivement l' ASCII (créé en 1963) en tant que système d'encodage sur l'internet.
2008-07: Le format PDF devient un standard ouvert et une norme ISO (ISO 32000-1:2008).
2008-11: La Commission européenne lance Europeana en tant que grande bibliothèque numérique publique européenne.
2010-03: Gallica, bibliothèque numérique de la Bibliothèque nationale de France (BnF), offre un million de documents tous supports confondus.
2010-06: Facebook fête ses 500 millions d'usagers.
2010-12: Une licence Creative Commons est utilisée par 400 millions d'oeuvres sur l'internet.
2011-01: Wikipédia fête ses dix ans avec 17 millions d'articles dans 270 langues.

2011-03: L’internet mondial compte 2 milliards d’usagers, soit 30,2% de la population mondiale.
2013: L’Ethnologue, répertoire encyclopédique de langues vivantes, publie pour la première fois sa version en ligne gratuite avant sa version imprimée payante.
2013-11: Creative Commons publie les versions 4.0 de ses licences.
2014-12: Une licence Creative Commons est utilisée par 882 millions d’oeuvres sur l’internet.
2015-04: The Online Books Page recense deux millions de livres en accès libre sur le web.
2015-05: Le web mondial compte près d’un milliard de sites.
2015-07: Les usagers de l’internet dont la langue maternelle n’est pas l’anglais atteignent les 75%.

Written by marielebert

2015-12-07 at 12:40

Posted in Uncategorized